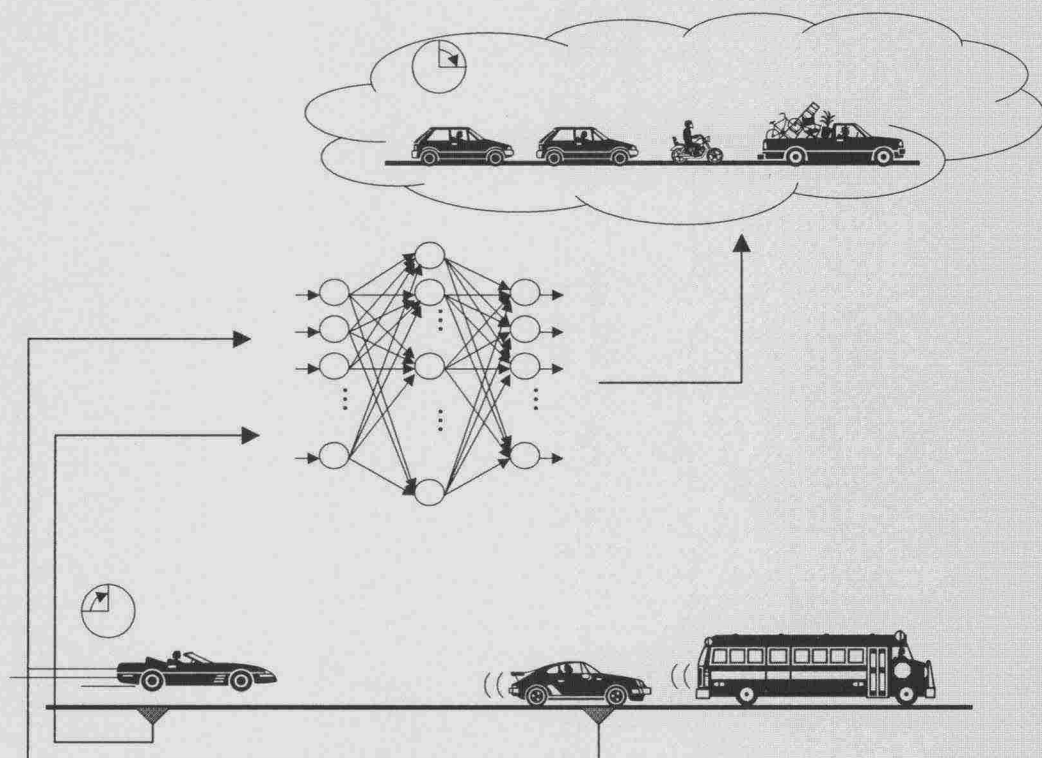


Liikennemäärän ja nopeuden lyhyen aikavälin ennustaminen



Tielaitoksen
selvityksiä

54/2000

Helsinki 2000

TIEHALLINTO
Liikenteen palvelut

Tielaitoksen selvityksiä
54/2000

Satu Innamaa, Matti Pursula

Liikennemäärän ja nopeuden lyhyen aikavälin ennustaminen

Tielaitos
TIEHALLINTO

Helsinki 2000

ISSN 0788-3722
ISBN 951-726-702-9
TIEL 3200640

Edita Oy
Helsinki 2000

Julkaisua saatavana:
Tiehallinto, julkaisumyynti
Faksi 0204 22 2652
s-posti julkaisumyynti@tiehallinto.fi
www.tielaitos.fi/kirjasto/tilaus.htm



TIEHALLINTO
Liikenteen palvelut
Opastinsilta 12 A
PL 33, 00521 HELSINKI
Puhelinvaihte 0204 22 150

Satu Innamaa, Matti Pursula: Liikennemäärän ja nopeuden lyhyen aikavälin ennustaminen. Helsinki 2000. Tielaitos, Tielaitoksen selvityksiä 54/2000. 101 s. + liitt. 3 s. ISBN 951-726-702-9. ISSN 0788-3722. TIEL 3200640.

Asiasanat: liikenne-ennusteet, ennusteet, telematiikka, liikenteen hallinta, liikenteen ohjaus, liikennevirta

Aiheluokka: 11, 20, 22

TIIVISTELMÄ

Dynaamiset liikenneinformaatio- ja liikenteenohjausjärjestelmät eivät voi toimia reaktiivisesti, vaan niiden täytyy osata ennakoida ja sopeuttaa strategiansa ajoissa muuttuviin liikennetilanteisiin. Tutkimuksen tarkoituksena oli selvittää eri tekijöiden vaikutuksia lyhyen aikavälin liikennetilanne-ennusteisiin. Ennusteet on tehty MLP-neuroverkkojen avulla. Ennustemalleja tehtiin ensisijaisesti Länsiväylälle.

Tutkimuksessa haluttiin selvittää, minkälainen on lyhyen aikavälin liikennetilanne-ennusteen kannalta optimaalinen neuroverkko. Tämän jälkeen optimoitiin muita ennustemalliin liittyviä tekijöitä, kuten syötteitä, ennustejakson pituutta ja poikkileikkausten sijaintia. Lopuksi neuroverkkoon perustuvia malleja verrattiin naiviin malliin ja Länsiväylällä kerättyjä kokemuksia sovellettiin kaksikaistaisessa kohteessa, valtatiellä 3 Valkeakoskella.

Parhaaksi aktivaatiofunktioyhdistelmäksi osoittautui piilokerroksen hyperbolinen tangentti yhdistettynä tulostuskerroksen lineaariseen funktioon. Syöteaineistossa annettujen liikennetietojen aikasarjan pituus osoittautui tärkeämmäksi kuin niiden poikkileikkauksien määrä, joilta tietoa kerätään.

Päivätyyppi- ja kellonaikatiedot eivät parantaneet ennusteita. Nopeuskeskihajontatiedot sitä vastoin paransivat alhaisimpien nopeuksien ennusteita. Puolen tunnin ennustemallin tulokset olivat kauttaaltaan hieman huonompia kuin 15 minuuttia ennustavan mallin. Kun poikkileikkausten määrä on rajallinen, on ole-massa etäisyys (tässä kohteessa 800 metriä), jota lähemmäksi lisätietopoikkileikkausta ei kannata sijoittaa ennustepoikkileikkauksesta.

Neuroverkkomalleille tehtiin vertailumalliksi malli, joka perustui liikenteen säännönmukaisuuteen. Tämä ns. naivi malli ei voi ennustaa alhaisia ruuhkanopeuksia, jollei ruuhka ole toistuvaa. Olennainen ero mallien välillä on se, että neuroverkkomalli on mahdollista opettaa reagoimaan satunnaisiin häiriötilanteisiin, toisin kuin naivi malli.

Kaksikaistaiselle tielle tehdyn ennustemallin perusteella havaittiin, että on tärkeämpää saada liikennetiedot kahdesta pisteestä kuin että tiedot saataisiin molemmilta suunnilta. Vastaantulevan liikenteen tietojen vaikutus mallin tekemän ennusteen hyvyyteen oli suhteellisen pieni. Tehdyn kokeilun perusteella vaikuttaisi siltä, että Länsiväylältä kerättyjä kokemuksia voidaan soveltaa suurelta osin suoraan myös kaksikaistaisilla teillä.

Satu Innamaa, Matti Pursula: Liikennemäärän ja nopeuden lyhyen aikavälin ennustaminen. [*Short-Term Prediction of Flow and Speed*]. Helsinki 2000. Finnish National Road Administration, Finnra Reports 54/2000. 101 p. + app. 3 p. ISBN 951-726-702-9. ISSN 0788-3722. TIEL 3200640.

Keywords: traffic forecasts, prediction, traffic telematics, traffic management, traffic control, traffic flow

ABSTRACT

Dynamic traffic information and control systems cannot function reactively but they should anticipate traffic situations and adjust their strategy to the changes in advance. The purpose of the research work was to study the influence of different factors to the short-term forecasts of the traffic situations. The forecasts were made with MLP neural networks. The prediction models were made primarily to Länsiväylä.

An objective of the research work was to find out the optimal structure of the neural network for the short term prediction of traffic situation. After this, other factors related to the prediction model were optimised, like inputs, the length of the prediction period, and the location of the cross-sections. Finally the models based on neural networks were compared to a naïve model and the experience gathered from Länsiväylä were applied to a two-lane environment, in the main road 3 in Valkeakoski.

The best combination of activation functions was a hyperbolic tangent for the hidden layer and a linear function for the output layer. The length of the time-series of the traffic information in the input turned out to be more important than the number of cross-sections where the data is collected.

The type of day and time of day information did not improve forecasts. However, the information about the standard deviation of the speed improved the forecasts of the lowest speeds. The results made by the model predicting 30 minutes ahead were throughout little worse than the results of the model predicting 15 minutes ahead. When the number of cross-sections is limited, there is a minimum for the distance (in this case 800 meters) between the prediction cross-section and the extra information cross-section.

A model based on the regular variations of the traffic was done as a comparison model to the neural network models. This so called naïve model cannot predict low rush hour speeds unless the congestion is recurrent. The essential difference between the models is that the neural network model can be trained to react to random incidents and non-recurrent congestion, unlike the naïve model.

Based on the model made for a two-lane road it could be discovered that it is more important to get the traffic information from two points than to get it from both directions. The impact of the traffic going to the opposite direction was relatively small to the performance of the model. Based on the experiment it seems that most of the experience gathered from Länsiväylä can be applied also in two-lane roads.

ALKUSANAT

Tielaitos sekä Liikenne- ja viestintäministeriö (LVM) ovat tilanneet tutkimuksen automaattisista liikenteenohjaus- ja liikenneinformaatiojärjestelmistä osana LVM:n TETRA – Liikenteen telematiikan rakenteiden tutkimus- ja kehittämisohjelmaa 1998 – 2000. Tämä tutkimus on valmistunut osana ko. toimeksiantoa.

Tutkimuksesta vastasi diplomi-insinööri Satu Innamaa Teknillisen korkeakoulun liikennelaboratoriosta ohjaajanaan professori Matti Pursula. Tutkimusapulaisena toimi tekniikan ylioppilas Kerkko Vanhanen.

Tutkimuksen etenemistä valvoi asiantuntijaryhmä, johon kuuluivat tekniikan tohtori Jussi Sauna-aho LVM:stä, diplomi-insinöörit Kari Hiltunen Tielaitoksen tiestötiedot-yksiköstä, Sami Luoma Tielaitoksen liikenteen palvelut -yksiköstä ja Pekka Rajala Uudenmaan tiepiiristä sekä tekniikan tohtorit Risto Kulmala VTT:ltä ja lisäksi Kosonen TKK:lta.

Neuraalilaskennassa on hyödynnetty CSC - Tieteellinen Laskenta Oy:n myöntämiä resursseja.

Sisältö

TIIVISTELMÄ	3
ABSTRACT	4
ALKUSANAT	5
1 Johdanto	9
2 Lyhyen aikavälin liikennetilanne-ennusteissa käytettyjä menetelmiä	10
2.1 Menetelmien välisiä vertailuja	10
2.2 Sijoittelu ja simulointi	11
2.3 Kalman-suodatus	12
2.4 Neuroverkot ja sumea päättely	14
2.5 Muita neuroverkkomalleja	14
3 Malli ja aineisto	16
3.1 Kohteet ja käytettävissä oleva aineisto	16
3.2 Valittu malli	18
4 MLP-verkko ja sen opettaminen	19
4.1 Monikerroserseptroniverkko	19
4.1.1 Yleistä	19
4.1.2 Aineiston muokkaus	20
4.1.3 Kerrosten ja neuronien määrät	21
4.1.4 Aktivaatiofunktiot	22
4.2 Vastavirta-algoritmiin perustuva opetus	22
5 Tutkimusasetelma ja hypoteesit	25
6 Optimaalinen neuroverkko	27
6.1 Perusennustemalli	27
6.2 Yksi malli vs. hajautettu malli	35
6.2.1 Oma malli kullekin ennustejaksolle	35
6.2.2 Erilliset liikennemäärä- ja nopeusmallit	38
6.2.3 Mallien välinen vertailu	42

7	Muiden tekijöiden optimoiminen	46
7.1	Päivä- ja kellonaikatiedot lisäsyötteinä	46
7.2	Nopeuskeskihajonta lisäsyötteenä	56
7.3	Ennustejakson pituus	61
7.4	Poikkileikkausten sijainti	67
7.4.1	Kaksi poikkileikkausta	67
7.4.2	Kolme poikkileikkausta	71
7.4.3	Lisätietopoikkileikkauksilla saavutettava etu	73
8	Naiivi vertailumalli	75
9	Ruuhkan ennustaminen	79
10	Sovellus kaksikaistaiselle tielle	85
10.1	Yleistä	85
10.2	Opetusalgoritmi	86
10.3	Liikennemäärä- ja nopeusennusteiden tulokset	86
10.4	Jonoprosenttiennusteiden tulokset	89
11	Johtopäätöksiä	92
12	Yhteenveto	96
	Lähdeluettelo	99
	Käytetyt symbolit	101

LIITTEET

Pääkomponenttianalyysi

1 Johdanto

Dynaamiset liikenneinformaatio- ja liikenteenohjausjärjestelmät eivät voi toimia reaktiivisesti, vaan niiden täytyy osata ennakoida ja sopeuttaa strategiansa ajoissa muuttuviin liikennetilanteisiin. Myös nykytilanteen tuntemiseen tarvitaan ennuste, koska liikennetiedot saadaan järjestelmän käyttöön aina pienellä viiveellä. Suomessa käytössä olevat järjestelmät eivät vielä tee ennusteita liikennetilanteen kehityksestä. Erityisesti ruuhkaan perustuvassa ohjauksessa tämä olisi kuitenkin tärkeää ja tästä syystä on tarpeen selvittää yksinkertaisia ennustamismenetelmiä.

Olemassa olevat ennustamismenetelmät on kehitetty pääosin matka-aikojen ja liikennemäärien ennustamiseen. Nämä parametrit, kuten myös liikennetiheys- ja pistenopeusennusteet, sopivat hyvin ruuhkavaroitus- ja liikenteenohjausjärjestelmien tarpeisiin.

Tämän tutkimuksen tarkoituksena oli selvittää eri tekijöiden vaikutuksia lyhyen aikavälin liikennetilanne-ennusteisiin. Ennusteet on tehty monikerrosperspektiivisillä neuroverkkojen (MLP-verkot) avulla. Jotta saataisiin tehtyä optimaalinen ennustemalli, täytyy etsiä paras neuroverkko ja optimaalinen tapa kerätä liikennedataa. Malliin liittyviä selvitettäviä asioita ovat mm. aineiston esikäsittelymenetelmä, neuroverkon aktivaatiofunktiot ja se, sovelletaanko yksittäistä vai hajautettua mallia. Lisäksi selvitetään ennustejakson pituuden ja mittausjärjestelyihin liittyvistä asioista poikkileikkausten määrän ja sijainnin vaikutusta tuloksiin.

Selvityksen luvussa 2 on esitelty lyhyen aikavälin liikenne-ennusteisiin käytettyjä malleja. Asiasta on kerrottu tarkemmin aiemmassa tutkimuksessa (Innamaa 1999). Luvussa 3 on esitelty valittu malli, ennusteiden kohde ja käytettävissä oleva data. Luvussa 4 on esitetty tarkemmin MLP-neuroverkko ja siihen liittyvät asiat. Hypoteesit sekä tutkimusasetelma on esitetty luvussa 5. Luvussa 6 on etsitty optimaalista neuroverkkoa ja luvussa 7 yritetty optimoida muita ennustemalliin liittyviä tekijöitä. Luvussa 9 on selvitetty, kuinka hyvin ruuhkatilanteiden dynaamisia muutoksia voidaan ennustaa. Luvussa 8 on esitelty vertailuksi tehty naiivi ennustemalli ja luvussa 10 on kokeiltu, kuinka Länsiväylällä saadut tulokset voidaan soveltaa kaksikaistaiselle tielle. Luvussa 11 on esitetty johtopäätöksiä. Liite 1 esittelee pääkomponenttianalyysin periaatteet.

2 Lyhyen aikavälin liikennetilanne-ennusteissa käytettyjä menetelmiä

2.1 Menetelmien välisiä vertailuja

Dynaamisten liikenneinformaatio- tai liikenteenohjausjärjestelmien täytyy osata ennakoida liikennetilanteet ja sopeuttaa strategiansa ajoissa. Se, että järjestelmä pystyy jatkuvasti päivittämään liikennemäärä- ja matka-aikaennusteita useita minuutteja eteenpäin käyttämällä reaaliaikaista liikennedatata, on ehdoton dynaamisen liikenteenohjauksen ja liikennetietojärjestelmien toiminnan edellytys. (Smith & Demetsky 1997.)

Smith & Demetsky (1997) tekemässä vertailussa tutkittiin kahta uutta liikenteen ennustamismenetelmää, jotka perustuivat neuroverkkoon ja ei-parametriseen regressiomalliin (lähin naapuri -menetelmä), sekä kahta perinteistä mallia, jotka perustuivat aikaisempaan keskiarvoon ja aikasarjoihin. Mallien tavoitteena oli ennustaa liikennemäärä seuraavalle 15 minuutille, kun tunnettiin liikennemäärätiedot viimeiseltä 15 minuutilta. *Taulukossa 1* on esitetty yhteenveto mallien heikkouksista ja vahvuuksista moottoriteliikenteen ennustamisessa.

Taulukko 1. Keskiarvo-, ARIMA-, neuroverkko- ja lähin naapuri -mallin välinen vertailu (Smith & Demetsky 1997).

Malli	Vahvuudet	Heikkoudet
Keskiarvo	<ul style="list-style-type: none"> ▪ Helppo ottaa käyttöön ▪ Nopea laskea 	<ul style="list-style-type: none"> ▪ Ei kykene vastaamaan odottamattomiin tapahtumiin
ARIMA	<ul style="list-style-type: none"> ▪ Aikasarjasovellus ▪ Taustalla vahva teoria 	<ul style="list-style-type: none"> ▪ Puuttuvat havainnot tuottavat ongelmia
Neuroverkko	<ul style="list-style-type: none"> ▪ Sopiva monimutkaisten, ei-lineaaristen riippuvaisuuksien kuvaamiseen 	<ul style="list-style-type: none"> ▪ Black box -malli ▪ Monimutkainen opettamisproseduuri
Lähin naapuri	<ul style="list-style-type: none"> ▪ Ei vaadi oletuksia olemassa olevista riippuvaisuuksista ▪ Hahmontunnistussovellukset 	<ul style="list-style-type: none"> ▪ Naapureiden löytämisen hankaluus

Smith & Demetsky (1997) vertailivat mallien hyvyttä neljän parametrin avulla: absoluuttisella virheellä, jakaumavirheellä, mallin käyttöönottamisen helpoudella ja mallin siirrettävyydellä. Lähin naapuri -menetelmä osoittautui paremmaksi kuin muut. Toiseksi parhaaksi tuli neuroverkko ja kolmanneksi jäi -keskiarvomalli. Vaikkei ARIMA-mallia voitu käyttää vajaalla datalla, se arvioitiin hie-man keskiarvomallia paremmaksi.

Parhaaksi osoittautuneen lähin naapuri -menetelmän tulosten absoluuttinen keskivirhe oli merkittävästi pienempi kuin muiden mallien. Keskiarvomallilla saatiin suurimmat absoluuttiset keskivirheet, vaikkei sillä ollutkaan taipumusta yli- tai aliarvioida ennusteitaan vahvasti. ARIMA-mallin absoluuttiset virheet olivat samaa suuruusluokkaa kuin muilla malleilla, mutta se yli- tai aliarvioi ennusteitaan usein huomattavasti. ARIMA-mallia ei kuitenkaan voitu käyttää vajaalla datalla,

mikä rajasi sen käyttöä huomattavasti. Neuroverkko osoittautui vertailussa toiseksi parhaaksi. Mallin ongelmana oli kuitenkin huomattava jakaumavirhe, jonka aiheuttaja oli todennäköisesti neuroverkon opetus. Neuroverkon siirrettävyys osoittautui huonoksi. (Smith & Demetsky 1997.)

Smith & Demetsky (1994) tekivät jo aikaisemmin vastaavanlaisen vertailun neuroverkko-, ARIMA- ja keskiarvomallin välillä. Tämän vertailun tulokset olivat hyvin samansuuntaisia. Keskiarvomalli toimi hyvin "tavallisina" päivinä, muttei kyennyt reagoimaan epätavallisiin tilanteisiin, jotka vaikuttivat liikennemääriin. ARIMA-mallin ennuste tuntui laahaavan jatkuvasti yhden tarkastelujakson jäljessä. Malli myös tuntui liioittelevan ennustetta voimakkaasti ylös- tai alaspäin. Neuroverkkomalli selviytyi ennustamisesta tarkasti, erityisesti ruuhkahuipun aikana, eikä ennuste kärsinyt ARIMA-mallin ongelmista.

Lee ym. (1998) ovat tehneet vertailun erilaisten matkanopeuksien lyhyen aikavälin ennustamismenetelmien välillä. He vertasivat usean selittäjän ensimmäisen asteen regressiomallia, ARIMA-mallia, vastavirta-algoritmillä opetettua neuroverkkomallia, jossa oli yksi piilokerros, ja Kalman-suodatinmallia. Heidän tuloksensa mukaan neuroverkko- ja Kalman-suodatinmalli tekivät tarkempia ennusteita kuin regressio- tai ARIMA-malli. Kalman-suodatinmalli oli jonkin verran parempi kuin neuroverkko, mutta neuroverkko antoi parempia tuloksia, kun malli siirrettiin alkuperäisestä kohteestaan muualle.

2.2 Sijoittelu ja simulointi

Hobeika & Ozbay (1991) kehittivät simulointiin perustuvan liikenteensijoittelumallin, jonka päätarkoitus oli testata erilaisia reittiohjausstrategioita, jotta liikenneverkko saataisiin toimimaan parhaalla mahdollisella tavalla. Mallin antamia tuloksia voitiin käyttää matka-aika- ja viivytysennusteina reaaliaikaisissa liikennetietojärjestelmissä.

Malli käytti iteratiivista prosessia ennustaessaan liikennemääriä ja jonoja sitä mukaa, kun ne muuttuivat. Malli perustui siihen, että simulointimallin tuloksena saatuja tietoja käytettiin hyödyksi liikenteen sijoittelussa, jolloin löydettiin nopein mahdollinen reitti kullekin lähtöpaikka-määräpaikkaparille verkolla. Menettely toistettiin, kunnes saavutettiin tasapaino. (Hobeika & Ozbay 1991.)

Myös Stephanedes ym. (1990a) käyttivät liikenteen sijoittelua ja simulointia ennustamiseen. Menetelmä perustui iteratiiviseen palautesilmukkaan sijoittelu- ja simulointivaiheen välillä. Sijoitteluvaihe jakoi matkat verkolle ja simulointivaihe tuotti yksityiskohtaista tietoa liikenneverkon toiminnasta. Silmukka lakkasi toimimasta, kun kahdella peräkkäisellä iterointikierröksellä tulokseksi saadut linkkien matka-ajat olivat riittävän lähellä toisiaan.

Stephanedes ym. (1990a) saivat tulokseksi, että useimmissa tapauksissa simulointi-/sijoittelumenetelmän avulla saadut tulokset olivat lähempänä havaittuja tuloksia kuin perinteisillä liikenteen sijoittelumenetelmillä saadut tulokset. Mene-

telmän moduulirakenne ja simulointivaihe olivat sen vahvuus, koska tutkittua liikenneverkkoa voitiin kuvata sille parhaiten sopivalla simulointi- ja sijoittelumallilla. Mallin antamat tulokset olivat sitä tarkempia mitä sopivammilla komponenteilla estimointia suoritettiin ja mitä tarkempi lähtöpaikka-määräpaikkamatriisi oli.

2.3 Kalman-suodatus

Rotterdamin liikenteenseurantajärjestelmää varten suunniteltiin ruuhkan ennustamisjärjestelmä, jonka tavoitteena oli ennustaa suunnittain erään tielinkin kokonaisliikennemäärä, ruuhkatilanne ja keskimääräinen matka-aika. Järjestelmän pää tavoitteena oli tuottaa ennusteita reaaliaikaisesti siten, että ennusteet olivat riittävän tarkkoja ja että järjestelmä kykeni tuottamaan ennusteita kaikissa olosuhteissa (satunnaisessa ruuhkassa, onnettomuuksien ja tietöiden aikana jne.). Järjestelmän oli myös kyettävä ottamaan ennusteissaan huomioon dynaamiset liikenteenohjaustoimenpiteet ja niiden vaikutukset. Tehtävään valittiin DYNA-DTMS-malli. Malli oli yhdistelmä tilastollisesta liikennemallista ja dynaamisesta sijoittelumallista. (Ben-Akiva ym. 1992.)

Erittäin lyhyen aikavälin (1 - 10 minuuttia) ennusteiden tekemiseen käytettiin tilastollista mallia. Kaksi pääsyitä mallin valitsemiseen olivat sen nopeus ja tulosten tarkkuus tämältyyppisissä ennusteissa. Näissä erittäin lyhyen aikavälin ennusteissa reitinvalinnalla ei ole suurta merkitystä, minkä takia tilastollisen mallin melko yksinkertainen kuljettajan reitinvalintaproseduuri ei vaikuttanut tuloksiin. Ennustamiseen valittiin Kalman-suodatin, koska se suodattaa liikennemittaukset reaaliajassa, tuntuu luonnolliselta valinnalta ennusteiden tuottamiseen, on joustava, kalibroi itse itsensä eikä siinä ole stationaarisuusoletuksia ja koska se pitää luonnostaan sisällään häiriöiden havainnoinnin. (Ben-Akiva ym. 1992.)

Kalman-suodattimen perusajatus on siinä, että haluttujen muuttujien arvot ennustetaan aiempaan tutkittavasta järjestelmästä olevaan informaatioon perustuen siihen hetkeen asti, kunnes uusi mittaus on käytettävissä. Kun uusi mittaus-tulos sitten saadaan käyttöön, muuttujan estimaattia korjataan mittauksen ja ennusteen väliseen erotukseen verrannollisena. (Karvonen ym. 1982.)

Tilastollista menetelmää täydentämään valittiin dynaaminen liikenteensijoittelumalli (DTA), jonka tehtävänä oli tehdä pidemmän aikavälin ennusteita. Tällaisissa ennusteissa kuljettajan reitinvalinnalla oli suurempi merkitys. DTA:ssa oli kuitenkin kaksi haittapuolta: mallit olivat hitaita ja ne tarvitsivat dynaamisen lähtöpaikka-määräpaikkamatriisin lähtötiedoikseen. (Ben-Akiva ym. 1992.)

Lopuksi ennusteet yhdistettiin siten, että saatu ennuste oli mahdollisimman tarkka. Tämä tehtiin laskemalla ennusteista painotettu keskiarvo siten, että painokertoimet vaihtelevat ennustejakson pituuden mukaan ja yhdistetyn ennusteen varianssi minimoitui. (Ben-Akiva ym. 1992.)

Stephanedes ym. (1990b, Stephanedes 1991, Kwon 1991) kehittivät moottori-tieliikenteen kysynnälle ja jakautumiselle ennustusalgoritmin, joka tunnisti liikennevirran ominaisuudet reaaliajassa ja jota voitiin käyttää dynaamisessa liikenteenohjauksessa ja ajoneuvojen opastamisessa. Myös heidän mallinsa käytti liikenteen jakauman ennustamisessa rekursiivista parametrien tunnistamista, joka perustui laajennettuun Kalman-suodattimeen.

Malli jaettiin kahteen osamalliin. Niistä ensimmäinen ennusti jokaista aikajaksoa kohti rampille saapuvien ajoneuvojen osuuden niistä ajoneuvoista, joiden olisi mahdollista tulla rampille. Toinen osamalli käytti tietokannan tietoja niiden ajoneuvojen määrän ennustamiseen, joiden olisi mahdollista tulla rampille. Mallin parametrejä päivitettiin jatkuvasti laajennetun Kalman-suodattimen avulla käyttämällä ennustemalleja havaintoyhtälöinä. (Stephanedes ym. 1990b, Stephanedes 1991, Kwon 1991.)

Mallin keskimääräinen liikenteenjakauman ennustevirhe vaihteli välillä 5,4 - 8,8 prosenttia. Keskimääräinen ennustevirhe niiden ajoneuvojen määrässä, joiden olisi ollut mahdollista tulla rampille, oli korkea: 6,1 - 13,4 prosenttia. Jälkimmäisen virheen suuruus saattaa johtua siitä, ettei ennusteessa käytetty hyväksi ylivirran liikennetietoja. (Stephanedes ym. 1990b, Stephanedes 1991, Kwon 1991.)

Ashok & Ben-Akiva (1993) kehittivät periaatteet reaaliaikaiseen aikariippuvaisten lähtöpaikka-määräpaikkamatriisien estimoimiseen ja ennustamiseen. Ongelma muotoiltiin Kalman-suodattimeksi, jossa tilavektori koostui eroista, joita lähtöpaikka-määräpaikkaparien välisissä liikennemäärissä oli edellisiin estimaatteihin verrattuna. Lähtöpaikka-määräpaikkaparien välisten liikennevirran vaihteluiden ennusteita muokattiin kunkin aikajakson lopulla saatujen linkkikohtaisten liikennelaskentojen tulosten perusteella. Tässä prosessissa käytettiin hyödyksi tietoja verkolla jo olevien ajoneuvojen matka-ajoista ja reitinvalinnasta. Sen lisäksi, että lähtöpaikka-määräpaikkamatriisit generoitiin kulloisellekin nykyhetkelle, mallin avulla voitiin tehdä ennusteita tulevasta ja päivittää aikaisempia matriiseja. Lähtöpaikka-määräpaikkamatriisin estimointimoduli päivitti ("suodatti") sijoitteluestimaatteja vertaamalla linkkien liikennelaskentojen tuloksia estimaattien liikennemäärien kanssa. Tämän jälkeen estimoinnit tehtiin seuraaville aikajaksoille ja prosessi jatkui. (Ashok & Ben-Akiva 1993.)

Ashok & Ben-Akiva (1993) havaitsivat, että suodatettu estimaatti oli huomattavasti lähempänä todellisia arvoja kuin vastaavat tietokannan tietoihin perustuvat estimaatit. Ennusteen laatu heikkeni progressiivisesti ja ennustetuilla estimaateilla oli taipumus konvergoida tietokannan tietoihin perustuvien estimaattien kanssa, kun ennusteaika piteni. Suodattimen toiminnan testaamiseksi he kokeilivat ennusteprosessia huonon tietokannan omaavalle datalle siten, että tehtiin useampi suodatus. He saivat tulokseksi, ettei suodatettujen estimaattien laatu ole erityisen herkkä tietokannan tietojen laadulle.

2.4 Neuroverkot ja sumea päättely

Matsui & Fujita (1998) ovat kehittäneet neuroverkkoihin ja sumeaan päättelyyn perustuvan matka-ajan ennustamismenetelmän moottoriteiden liikenneinformaatiojärjestelmien käyttöön. Mallissa käytettiin ennusteen syötetietoina matka-ajaa ajoneuvon lähtöhetkellä ja tiejaksolla olevien ajoneuvojen lukumäärää.

Malliin tehtiin sumeaa päättelyä varten useita ohjaussääntöjä. Säännöt esitettiin numeerisessa muodossa neuroverkkoja varten. Myös raja-arvot määriteltiin neuroverkon avulla. Jos-osan selkeillä joukoilla oli sumean joukon piirteitä, koska ne korvasivat sumean logiikan jäsenyysfunktion neuroverkolla. Säännön sitten-osa koostui myös neuroverkosta.

2.5 Muita neuroverkkomalleja

Kwon & Stephanedes (1994) rakensivat neuroverkkomallin liikennemäärien ennustamiseen moottoritien poistumisrampeilla 5 minuutin jaksoissa. Ennustemalli rakennettiin kolmikerroksisena vastavirta-algoritmiin perustuvana neuroverkkona. Mallissa oli syöte- ja vastekerrosten välissä yksi piilokerros. Mallia rakennettaessa oletettiin, että poistumisrampin liikennemäärä oli riippuvainen sekä ylä- että alavirran liikenneolosuhteista. Malli opetettiin sekä tarkastelupäivän että aikaisempien päivien liikennemäärädatalla, jolloin neuroverkko antoi vasteena halutun liikennemäärän. (Kwon & Stephanedes 1994.)

Kwon & Stephanedes (1994) vertasivat neuroverkkomallia 1) olemassa olevaan UTCS-2-malliin, joka on erikseen kalibroitu kiinteäparametrinen, aikaisemmista tiedoista koottua tietokantaa ja nykyistä aineistoa käyttävä malli, ja 2) mukautuvaparametriseen malliin, joka ennusti liikennemäärät Kalman-suodattimen avulla. Jälkimmäinen vertailumalli koostui kahdesta alamallista: ensimmäinen oli suunniteltu normaaleihin liikenneolosuhteisiin ja se käytti hyväkseen sekä tietokantaa että nykyistä aineistoa, malleista toinen oli suunniteltu poikkeuksellisiin tilanteisiin, joissa tietokannan tietoihin perustuva kysyntämalli erosi huomattavasti tarkasteluhetkellä mitatusta, ja se käytti ainoastaan nykyistä aineistoa. Mallin parametrit oli sovitettu neuroverkon opetusdatan avulla.

Vertailussa mallien tekemille ennusteille laskettiin keskimääräinen absoluuttinen virhe ja keskineliövirhe. UTCS-2-malli oli jokaisessa vertailussa huonoin. Mukautuvaparametrinen malli ja neuroverkkomalli tekivät vuorotellen pienimmät virheet. Neuroverkon keskimääräinen absoluuttinen virhe ja keskineliövirhe olivat kuitenkin hieman pienempiä kuin mukautuvaparametrisella mallilla. (Kwon & Stephanedes 1994.)

Myös Zhang ym. (1997) kehittivät neuroverkkoon perustuvan ennustemallin dynaamisten liikenteenohjausjärjestelmien tarpeisiin. Neuroverkko rakennettiin mallintamaan Papageorgioun laajentamaa Paynen kontinuumimallia (Papageorgiou ym. 1990). Ennustamiseen käytetty neuroverkko perustui vastavirta-algoritmiin ja se koostui syöte- ja vastekerrosten lisäksi kahdesta piilokerrok-

sesta. Mallin syötteet ja vasteet valittiin käytetyn liikennemallin mukaisesti. Syötteinä olivat tarkasteluhetken nopeus tarkasteltavalla tiejaksolla ja sitä edeltävällä tiejaksolla, liikennetiheys tarkasteltavalla tiejaksolla ja sitä seuraavalla tiejaksolla ja rampille tulevien ajoneuvojen liikennemäärä ja vasteena nopeus tarkasteltavalla tiejaksolla ennustehetkellä.

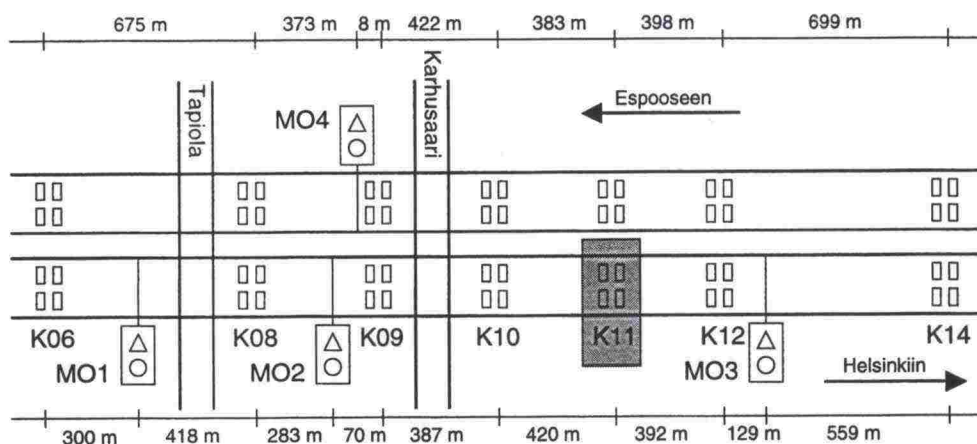
Mallia opetettaessa syötteet annettiin sekä liikenne- että neuroverkkomallille. Mallien antamia vasteita verrattiin ja erotus syötettiin takaisin neuroverkkoon, jolloin verkko korjasi painokertoimia ja kynnysarvoja omien oppimissääntöjensä mukaisesti. Näin neuroverkko oppi mallintamaan liikennettä valitun liikennemallin mukaisesti. Malli opetettiin simuloinnin avulla saadulla datalla. (Zhang ym. 1997.)

3 Malli ja aineisto

3.1 Kohteet ja käytettävissä oleva aineisto

Ennustemalleja tehtiin ensisijaisesti Länsiväylälle Espoon ja Helsingin rajalle. Länsiväylälle kehitettyjä menetelmiä testattiin kuitenkin myös kaksikaistaisessa kohteessa, valtatiellä 3 Valkeakoskella.

Länsiväylän ennustemallit kehitettiin Länsiväylän ruuhkaohjausjärjestelmän poikkileikkaukselle 11 Helsingin suuntaan (kuva 1). Kyseisen poikkileikkauksen lisäksi mallien tekemisessä voitiin käyttää hyväksi liikennetietoja kuudelta muulta poikkileikkaukselta. Näistä neljä sijaitsee ennen ko. poikkileikkausta ja kaksi sen jälkeen. Länsiväylän aineisto oli ajalta 1.5. – 3.6.1999, jolloin muuttuvat nopeusrajoitukset ja varoitusmerkit eivät olleet vielä käytössä tällä osuudella.



Kuva 1. Ennustemallit kehitettiin ennustamaan liikennettä Länsiväylän ruuhkavaroitussjärjestelmän poikkileikkauksessa 11.

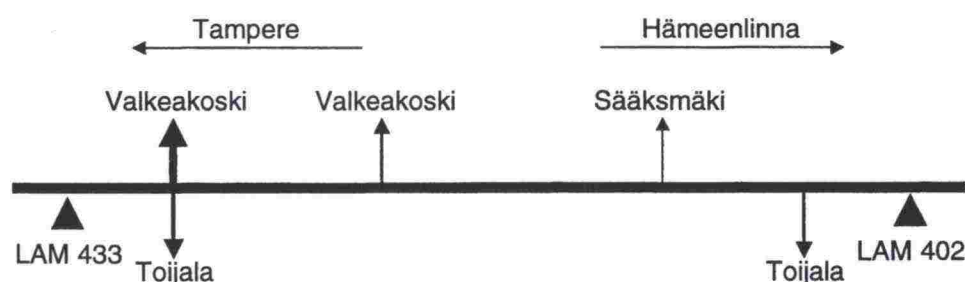
Länsiväylän aineistossa oli ongelmana induktioilmaisimista saadun aineiston huono laatu. Saaduista ilmaisuista oli virheellisiä keskimäärin 38 – 48 prosenttia (taulukko 2). Virheilmaisista selvästi yleisin oli virhe 6, joka tarkoitti sitä, että ajoneuvon liike oli epätasainen tai poikkeava, jolloin ajoneuvon pituus ei ollut mitattavissa. Käytännössä tämä tarkoittaa voimakkaasti kiihdyttynyttä tai jarrutettua ajoneuvoa. Tehdyissä ennustemalleissa oli kuitenkin tavoitteena ennustaa järjestelmän mittaama liikennemäärä ja keskinopeus - ottamatta kantaa arvojen oikeellisuuteen.

Taulukko 2. Havaintojen määrä ja virheellisten ilmaisujen osuus prosentteina poikkileikkauksittain Länsiväylällä ajalta 1.5. – 3.6.1999. Havaintojen lukumäärä (N) ja keskimääräinen virheellisten havaintojen osuus (e%) on laskettu molemmille suunnille yhteensä. Pienin ja suurin virheellisten havaintojen osuus ($e_{min}\%$, $e_{max}\%$) on laskettu päivittäisistä havainnoista.

	K06	K08	K09	K10	K11	K12	K14
e%	47	39	48	47	38	45	47
$e_{min}\%$	46	37	47	46	34	43	42
$e_{max}\%$	48	40	50	50	44	48	49
N	44 140	40 557	32 449	57 095	57 450	53 881	56 977

Tutkimus perustuu viiden minuutin liikennemäärä- ja nopeushavaintoihin. Havaintomatriisia rakennettaessa yksittäiset puuttuvat havainnot (kyseisen viisiminuuttisen aikana ei ollut riittävän monta "virheetöntä" havaintoa) ekstrapoloitiin edeltävästä ja seuraavasta havainnosta lineaarisesti. Mikäli puuttuvia havaintoja oli enemmän kuin yksi peräkkäin, ekstrapolointiin ei ryhdytty. Keskinopeus laskettiin ainoastaan "virheettömien" havaintojen perusteella, mutta liikennemäärä-estimaatissa olivat mukana kaikki havainnot. Aineistoa oli käytettävissä sekä poikkileikkaus- että kaistakohtaisena.

Valtatiellä 3 Valkeakoskella ennusteet tehtiin Pispantallissa sijaitsevan liikenteen automaattisen mittauspisteen (LAM-pisteen) 433 suunnalle 1 (Tampere) (kuva 2). Ennustetta varten saatiin lisätietoja myös LAM-pisteestä 402, joka sijaitsee Jutikkalassa. LAM-pisteiden välinen etäisyys oli 17,6 kilometriä.



Kuva 2. Valtatien 3 jakso, jolla ennustemallin tekemistä kokeiltiin kaksi-kaistaiselle tielle. Ennustemallit tehtiin LAM-pisteelle 433 ja malli sai lisätietoja LAM-pisteestä 402.

Ennustemallin tekoa varten valtatieltä 3 oli käytettävissä LAM-pisteiden ajoneuvokohtaiset liikennetiedot touko-lokakuulta 1999. Ajoneuvokohtainen aineisto muutettiin viiden minuutin liikennemäärä- ja keskinopeushavainnoiksi. Mallin tekoon käytetyistä havainnoista poistettiin ne viisiminuuttiset, joiden aikana yksikään ajoneuvo ei ollut ohittanut LAM-pistettä.

3.2 Valittu malli

Sijoitteluun ja simulointiin perustuvat ennustemalli voi olla varteenotettava vaihtoehto kohteessa, jossa verkkonäkökulma on tärkeä. Koska nyt kehitettävän mallin kohteena oli yksittäinen tie, jolla ei juurikaan ole vaihtoehtoisia reittejä, muut menetelmät vaikuttavat lupaavammilta vaihtoehdoilta.

Edellä kuvatuissa vertailututkimuksissa neuroverkko menestyi erittäin hyvin ja neuroverkkomallit olivat parempia kuin esim. ARIMA-mallit (Smith & Demetsky 1997 ja 1994, Lee ym. 1998). Neuroverkko oli erityisen hyvä ruuhkan ennustamisessa, josta perinteisemmät ennustemallit suoriutuivat huonosti. Tutkimuksissa neuroverkkojen ongelmat liittyivät opettamiseen tai siihen, että neuroverkko oli siirretty alkuperäisestä paikasta toiseen.

Neuroverkon ominaisuuksiin kuuluu se, että se oppii kunkin ongelman omat erityispiirteet, ja näin ollen ajatus neuroverkkoon perustuvan ennustemallin siirtämisestä alkuperäisestä paikastaan toiseen, ilman uutta opettamista, ei ole järkevä. Opettamiseen täytyy käyttää aikaa ja se täytyy tehdä huolella. Kalman-suodatuksen perustuvat mallit menestyivät vertailuissa yhtä hyvin kuin neuroverkkomallit. Kalman-suodatinmalli tarvitsee kuitenkin taustalleen jonkin muun mallin (Ben-Akiva ym. 1992), minkä takia nyt tehdyssä tutkimuksessa päädyttiin ennustemenetelmäksi valitsemaan monikerrosperspektiivi-neuroverkko (MLP-verkko). MLP-verkon etuna on myös sen käytön yksinkertaisuus, kun verkko on ensin saatu opetettua.

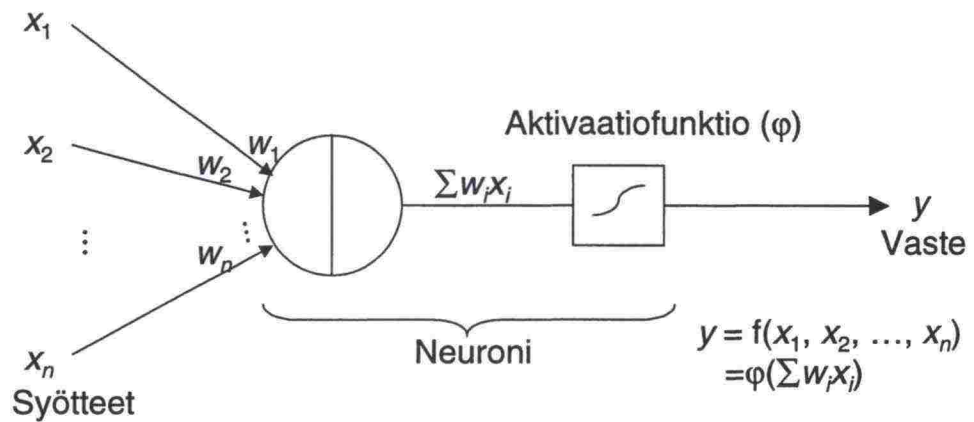
Neuroverkkomallia on tässä työssä verrattu naiiviin malliin, joka perustuu liikenteen säännöllisiin aikavaihteluihin.

4 MLP-verkko ja sen opettaminen

4.1 Monikerrosperseptroniverkko

4.1.1 Yleistä

Perseptronilla tarkoitetaan *kuvan 3* mukaista neuronin, joka jakaa syöteavaruuden kahteen osaan (luokkiin 1 ja 2). Monikerrosperseptroniverkko (multi-layer perceptron network, MLP) on eteenpäin syöttävä neuroverkko (feedforward network), joka koostuu useammasta kerroksesta perseptroneja. Eteenpäin syöttävä verkko koostuu kahdesta tai useammasta keskinäisesti toisensa poissulkevista neuronien tai kerrosten muodostamasta joukosta. (Schalkoff 1992.)

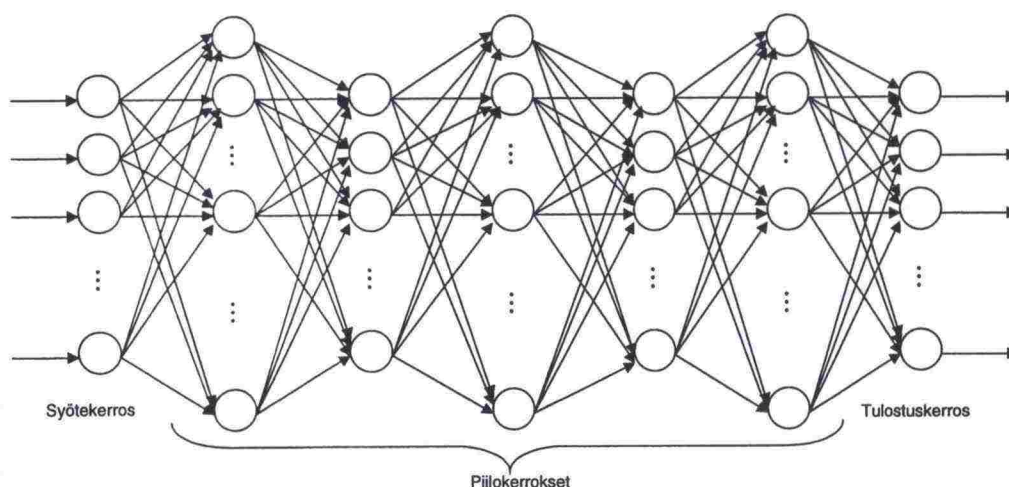


Kuva 3. Perseptroni.

Kerroksista ensimmäinen (syötekerros) ottaa vastaan verkolle annettavat syötteet ja kerroksista viimeinen (vaste- tai tulostuskerros) antaa verkon vasteen. Näiden kahden äärimmäisen kerroksen välissä on nolla, yksi tai useampia piilokerroksia (*kuvaa 4*). Linkit (tai painot, w) yhdistävät kunkin kerroksen neuronin kaikkiin seuraavaan (mutta ainoastaan seuraavaan) kerroksen neuroneihin. Yhteys on yksisuuntainen (vrt. eteenpäin syöttävä) (Schalkoff 1992). Neuronin välittää eteenpäin seuraavan neuronikerroksen neuroneille sen arvon y_j , jonka aktivaatiofunktio ϕ saa neuronin syötteiden y_i painotetulla summalla.

$$y_j = \phi_j \left(\sum_{i=0}^m w_{ji} y_i \right)$$

Yhtälöissä w_{ji} on neuronien i ja j välinen painokerroin ja m syötteiden määrä. Aktivaatiofunktioilla tarkoitetaan funktiota, joka rajoittaa vasteen arvoa (Haykin 1999).



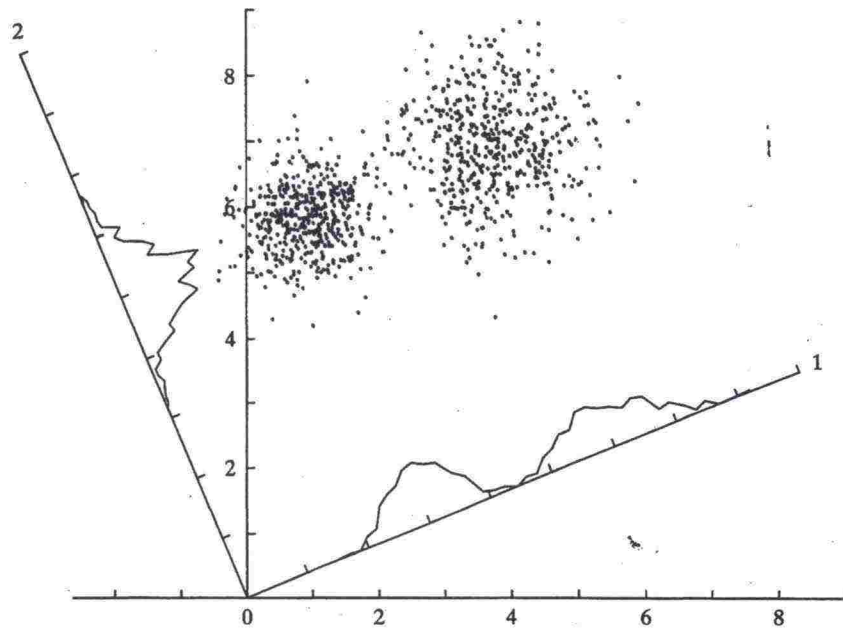
Kuva 4. Monikerroksinen eteenpäin syöttävä neuroverkko.

4.1.2 Aineiston muokkaus

Kun neuroverkkoa opetetaan aineistolla, jossa parametreille tyypilliset arvot ovat lukuarvoltaan hyvin erilaisia, aineisto on hyvä skaalata tai normeerata. Näin neuroverkon opetusta saadaan tehostettua (Demuth & Beale 1998). Skaalaamisella tarkoitetaan sitä, että kunkin parametrin arvoja muutetaan siten, että ne vaihtelevat vakiovälillä, kuten $[0, 1]$ tai $[-1, 1]$. Normeerauksessa aineisto muokataan siten, että sillä on vakiokeskiarvo ja -keskihajonta (esim. keskiarvona nolla ja keskihajontana yksi).

Joskus syöteparametrien määrä on suuri, mutta syötevektorien komponentit korreloivat keskenään. Tällaisissa tilanteissa on hyvä vähentää syöteavaruuden dimensiota (Demuth & Beale 1998). Jos vektori yksinkertaisesti katkaistaan, neuroverkon tekemä keskimääräinen neliövirhe tulee yhtä suureksi kuin katkaistujen elementtien varianssien summa. Olisi siis parempi löytää sellainen muunnos, joka olisi optimaalinen keskimääräinen neliövirhe -mielessä. Pääkomponenttianalyysi on eräs tällainen menetelmä (Haykin 1999).

Pääkomponenttianalyysissä syötevektorien komponentit muokataan kohtisuoriksi siten, etteivät ne enää korreloi keskenään. Kohtisuorat komponentit järjestetään variaatiokertoimen mukaiseen suuruusjärjestykseen (ensimmäiseksi se, jolla on suurin kerroin). Lopuksi karsitaan ne komponentit, jotka selittävät vähiten aineiston vaihtelusta (Demuth & Beale 1998). Pääkomponenttianalyysiä on esitelty tarkemmin liitteessä 1. Kuvassa 5 on esimerkki pääkomponenttianalyysistä kahdessa dimensiossa. On hyvä muistaa, että vaikka tässä yksinkertaisessa esimerkissä aineiston klusterit on helppo erottaa toisistaan ilman pääkomponenttianalyysiäkin, tilanne ei aina ole yhtä selvä moniulotteisemmissa käytännön ongelmissa.



Kuva 5. Kaksiulotteinen pistejoukko ja sen tiheysfunktiot, jotka projisioituvat akselleille 1 ja 2. Akselin 1 projektiossa varianssi on maksimissaan ja tämän akselin tiheysfunktiosta voidaan selvästi erottaa aineiston kaksijakoisuus. (Haykin 1999.)

4.1.3 Kerrosten ja neuronien määrät

Neuroverkkoa tehtäessä pitää ensin päättää neuronikerrosten ja niissä olevien neuronien määrät. Syötekerroksia on aina yksi ja sen neuronien määrä on sama kuin syöteparametrien lukumäärä. Vastaavasti tulostuskerroksia on yksi ja se koostuu yhtä monesta neuronista kuin verkolla on vasteparametrejä. Piilokerroksia sen sijaan voi olla nolla tai useampia. Piilokerros yhdistää edeltävän ja seuraavan kerroksen syötteet ja vasteet uudelleen siten, että aineisto luokituu paremmin. Piilokerroksilla voidaan myös tuoda esille verkon eri syötteiden väliset yhteydet (Schalkoff 1992).

Jos järjestelmä on niin kutsuttu musta laatikko (tunnetaan syötteet ja vasteet, muttei sitä mitä järjestelmässä tapahtuu), siihen kannattaa yleensä laittaa yksi piilokerros. Useampaa piilokerrosta voi perustella, jos järjestelmän toimintaa tunnetaan (Oja 1998). Nyt tehdyssä tutkimuksessa piilokerrosten määräksi valittiin aina yksi.

Piilokerroksen neuronien lukumäärän yläraja saadaan määriteltä Widrow:n säännön avulla, jonka mukaan opetusjoukon koon tulisi olla vähintään kymmenen kertaa verkon painokertoimien (eli estimoitavien parametrien) lukumäärä. Jos meillä on T opetusvektoria, saadaan piilokerroksen neuronien maksimimäärä (N_h) ratkaistua syöte- ja tulostuskerroksen neuronien (N_i ja N_o) määrän avulla. (Oja 1998.)

$$(N_i + 1)N_h + (N_h + 1)N_o = \frac{1}{10}T$$

$$\Rightarrow N_h = \frac{\frac{1}{10}T - N_o}{N_i + N_o + 1}$$

4.1.4 Aktivaatiofunktiot

Yksinkertaisin aktivaatiofunktio on porraskunktio, joka voi olla esimerkiksi muotoa

$$\varphi(u) = \begin{cases} +1 & \text{jos } u \geq 0 \\ -1 & \text{jos } u < 0 \end{cases}$$

Yksinkertaisuudestaan huolimatta porraskunktio ei sovi moniin sovelluksiin differentioitumattomuutensa takia. Sen sijaan käytetään yleensä hieman monimutkaisempia aktivaatiofunktioita, joilla on etuna differentioituvuus. Tällaisia ovat esimerkiksi sigmoidinen tai lineaarinen funktio.

Tyypillisiä esimerkkejä sigmoidisista funktioista ovat logistinen funktio tai hyperbolinen tangenttifunktio. Logistisen funktion arvot liikkuvat välillä $[0, 1]$ ja se on muotoa

$$\varphi(u) = \frac{1}{1 + e^{-\beta u}}$$

Hyperbolinen tangenttifunktio taas saa arvot väliltä $[-1, 1]$ ja se on muotoa

$$\varphi(u) = \tanh(\beta u) = \frac{e^{\beta u} - e^{-\beta u}}{e^{\beta u} + e^{-\beta u}}$$

Kerroin β kuvaa sigmoidisen funktion jyrkkyyttä origossa. Jos β on hyvin suuri, molemmat funktiot lähestyvät porraskunktiota. (Oja 1998.)

4.2 Vastavirta-algoritmiin perustuva opetus

MLP-verkko voidaan opettaa usealla eri menetelmällä, joista yksi on vastavirta-algoritmi (back-propagation algorithm). Neuroverkon tulostuskerroksessa olevan neuronin j vasteen virhe n :nnellä opetuskierroksella (n :nnen opetusvektorin jälkeen) $e_j(n)$ on määritelty toivotun vasteen $d_j(n)$ ja verkon antaman vasteen $y_j(n)$ välisenä erotuksena.

$$e_j(n) = d_j(n) - y_j(n)$$

Neuronin j hetkellinen virhe-energian määrä on yhtä suuri kuin $\frac{1}{2}e_j^2(n)$. Vastavasti kokonaisenergian hetkellinen määrä $E(n)$ saadaan summaamalla $\frac{1}{2}e_j^2(n)$ -termit yli kaikkien tulostuskerroksen neuronien.

$$E(n) = \frac{1}{2} \sum_{j \in C} e_j^2(n)$$

Yhtälössä joukko C pitää sisällään kaikki tulostuskerroksen neuronit. Merkitään T :llä opetusjoukon opetusvektoreiden määrää. Keskimääräinen neliövirheen energia saadaan summaamalla $E(n)$ yli $n:n$ ja normeeraamalla se joukon koon T suhteen.

$$E_{av} = \frac{1}{T} \sum_{n=1}^T E(n)$$

Hetkellinen virhe-energia, ja näin ollen myös keskimääräinen virhe-energia, on kaikkien verkon vapaiden parametrien funktio. Tietylle opetusjoukolle E_{av} edustaa kustannusfunktioita, jonka avulla kuvataan opetuksen etenemistä. Oppimisprosessin tavoitteena on säätää verkon vapaat parametrit siten, että E_{av} minimoituu. (Hayden 1999.)

Kuvitellaan monikerrosperserptroniverkko. Kun neuronista j menee aktivaatiofunktioon ϕ signaali $v_j(n)$, neuronin vasteena on $y_j(n)$.

$$v_j(n) = \sum_{i=0}^m w_{ji}(n) y_i(n)$$

$$y_j(n) = \phi_j(v_j(n))$$

Kullakin opetuskierroksella painoja $w_{ji}(n)$ korjataan $\Delta w_{ji}(n)$:n verran, joka on verrannollinen osittaisdifferentiaaliin $\partial E(n)/\partial w_{ji}(n)$. Ketjusäännön avulla gradientti saadaan seuraavanlaiseen muotoon.

$$\frac{\partial E(n)}{\partial w_{ji}(n)} = \frac{\partial E(n)}{\partial e_j(n)} \cdot \frac{\partial e_j(n)}{\partial y_j(n)} \cdot \frac{\partial y_j(n)}{\partial v_j(n)} \cdot \frac{\partial v_j(n)}{\partial w_{ji}(n)}$$

Osittaisdifferentiaali $\partial E(n)/\partial w_{ji}(n)$ on herkkyystekijä, joka määrää optimiratkaisun hakusuunnan painoavaruudessa. Yhtälön ratkaisu saadaan ratkaisemalla osa-ongelmat.

$$\frac{\partial E(n)}{\partial w_{ji}(n)} = e_j(n) \cdot (-1) \cdot \phi_j'(v_j(n)) \cdot y_i(n) = -e_j(n) \phi_j'(v_j(n)) y_i(n)$$

Korjaustermin $\Delta w_{ji}(n)$ suuruus määritellään delta-säännön avulla.

$$\Delta w_{ji}(n) = -\eta \frac{\partial E(n)}{\partial w_{ji}(n)}$$

Yhtälössä η on vastavirta-algoritmin oppimisaste (learning rate). Negatiivinen etumerkki selittyy gradientin pienenemisellä painoavaruudessa. (Haykin 1999.)

Perusvastavirta-algoritmi on siis gradienttimenetelmä. Algoritmista on kuitenkin olemassa lukuisia sovelluksia, jotka konvergoivat nopeammin kuin perusgradi-

enttimenetelmä. Näitä ovat esimerkiksi konjugaattigradienttimenetelmä, Newtonin menetelmä ja siitä johdettu Levenberg-Marquardt'in menetelmä. (Demuth & Beale 1998.)

Newtonin menetelmässä parametrien (\mathbf{X}) päivitys tehdään seuraavan yhtälön mukaisesti.

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{H}_k^{-1} \mathbf{g}_k$$

Yhtälössä \mathbf{H}_k on hyvyysindeksin (performance) Hessen matriisi (toiset derivaatat) nykyisillä painojen ja harhakertoimien arvoilla ja \mathbf{g} gradientti. Newtonin menetelmää käytetään usein vaihtoehtona konjugaattigradienttimenetelmälle, kun neuroverkko halutaan opettaa tavallista gradienttimenetelmää nopeammin. Newtonin menetelmä konvergoi usein nopeammin kuin konjugaattigradienttimenetelmä, mutta valitettavasti Hessen matriisin laskeminen eteenpäin syöttävälle neuroverkolle on usein erittäin työlästä.

Levenberg-Marquardt-algoritmia suunniteltaessa tavoitteena oli löytää sellainen perusgradienttimenetelmä nopeampi opetusalgoritmi, joka ei vaadi Hessen matriisin laskemista. Kun hyvyysfunktio on neliösummamuodossa, Hessen matriisi \mathbf{H} voidaan approksimoida seuraavalla tavalla.

$$\mathbf{H} \approx \mathbf{J}^T \mathbf{J}$$

Nyt gradientti \mathbf{g} voidaan laskea seuraavan yhtälön mukaisesti.

$$\mathbf{g} = \mathbf{J}^T \mathbf{e}$$

Yhtälöissä \mathbf{J} on Jacobin matriisi, joka koostuu verkon tekemien virheiden ensimmäisistä derivaatoista suhteessa painoihin ja harhatermeihin ja \mathbf{e} vektori, joka koostuu verkon virheistä. Jacobin matriisi voidaan laskea perusvastavirta-algoritmeilla, mikä on huomattavan paljon yksinkertaisempaa kuin Hessen matriisin laskeminen. (Demuth & Beale 1998.)

Levenberg-Marquardt-algoritmi käyttää hyväkseen tätä Hessen matriisin approksimaatiota, jolloin parametrejä päivitetään seuraavan yhtälön mukaisesti.

$$\mathbf{x}_{k+1} = \mathbf{x}_k - [\mathbf{J}^T \mathbf{J} + \mu \mathbf{I}]^{-1} \mathbf{J}^T \mathbf{e}$$

Jos skalaari μ on nolla, saadaan tulokseksi Newtonin menetelmä Hessen matriisin approksimaatiolla. Kun μ on suuri, tuloksena on gradienttimenetelmä pienellä askelkoolla. Newtonin menetelmä on nopeampi ja tarkempi virheen minimin läheisyydessä, minkä takia tavoitteena on siirtyä gradienttimenetelmästä Newtonin menetelmään mahdollisimman nopeasti. Tästä syystä μ pienenee jokaisen onnistuneen askeleen (hyvyysfunktion arvo laskee) jälkeen ja sitä kasvatetaan ainoastaan silloin, kun askel muuten kasvattaisi hyvyysfunktion arvoa. Tällä tavoin hyvyysfunktion arvo pienenee jokaisella iteraatiokierroksella. (Demuth & Beale 1998.)

5 Tutkimusasetelma ja hypoteesit

Ennustepoikkileikkauksen liikennetilanne on saapuvan liikennemäärän, ruuhkan sijainnin, tien välityskyvyn ym. ominaisuuksien lisäksi riippuvainen vallitsevista keli- ja sääolosuhteista. Myös aggregointijakson pituudella on merkitystä, sillä pitkä jakso tasaa vaihteluita. Se taas, mitä lisäarvoa muiden poikkileikkausten liikennetiedoista on syötetietoina mallille, on riippuvainen näiden lisätietopoikkileikkausten sijainnista suhteessa ruuhkapisteisiin tai ramppeihin ja siitä, mikä on ennustejakson pituus.

Ennustemallia tehtäessä tulisi periaatteessa selvittää kaikkien em. tekijöiden vaikutukset malliin sen lisäksi, että on tarpeen etsiä optimaalinen mallinnusmenetelmä (neuroverkko tms.). Tässä selvityksessä sää- ja kelitiedot piti jättää tutkimuksen ulkopuolelle, koska aineistossa ei ollut suuria sää- tai kelivaihteluita (aineisto kerättiin touko-kesäkuussa). Myöskin lyhyempien aggregointi- ja ennustejaksojen vaikutus jää selvittämättä aineiston ongelmien vuoksi (mm. havaintojen määrä ei riitä minuutin aggregointijaksoon). Lisätietopoikkileikkausten optimaalisen sijainnin määrittämisessä rajoittavana tekijänä oli koealueen lyhyys (noin kolme kilometriä). Muiden tekijöiden vaikutuksia sitä vastoin voitiin selvittää ainakin jossain laajuudessa.

Tutkimuksessa haluttiin ensin selvittää, minkälainen on lyhyen aikavälin liikennetilanne-ennusteen kannalta optimaalinen neuroverkko. Tällä tarkoitetaan sitä, mikä aineiston esikäsittelymenetelmä kannattaisi valita, minkälaiset aktivaatiofunktiot laittaa neuroverkkoon tai että kannattaako malli jakaa osamalleihin. Tämän jälkeen optimoitiin muita ennustemalliin liittyviä tekijöitä, kuten syötteitä, ennustejakson pituutta ja poikkileikkausten sijaintia. Lopuksi neuroverkkoon perustuvia malleja verrattiin naiiviin malliin ja Länsiväylällä kerättyjä kokemuksia sovellettiin kaksikaistaisessa kohteessa valtatiellä 3.

Optimaalinen neuroverkko yritettiin siis löytää selvittämällä optimaalinen aineiston esikäsittelymenetelmä, neuroverkon aktivaatiofunktioyhdistelmä ja osamallijako. Erilaisia vaihtoehtoja kokeiltiin hieman erilaisilla aineistoilla, jotta nähtiin ominaisuuksien vahvuudet erilaisissa tilanteissa. Aineiston esikäsittelymenetelmävaihtoehtoiksi valittiin skaalaus ja normeeraus yhdistettynä pääkomponenttianalyysiin. Aktivaatiofunktiovaihtoehtoina käytettiin lineaarista funktiota, logistista funktiota ja hyperbolista tangenttia. Osamalleihin jakamisen kannattavuutta selvitettiin vertaamalla yksittäistä mallia, joka teki kaikki ennusteet, osamalleihin jaettuihin malleihin, joissa yksittäinen osamalli teki ennusteen joko yksittäiselle ennustejaksolle tai suurelle. Alla ovat asiaan liittyvät tutkimushypoteesit.

- Syötedatan dimensioiden vähentäminen pääkomponenttiallyysin avulla kannattaa silloin, kun syöteparametrejä on paljon.
- Aktivaatiofunktioyhdistelmien välillä ei ole eroa ennusteen tarkkuuden kannalta.

- Osamalleihin jaetun ennustemallin tulokset ovat vähintään yhtä hyviä kuin jakamattoman mallin.

Ennustemallia yritettiin lisäksi optimoida tutkimalla erilaisia syötteitä, ennustajakson pituutta ja poikkileikkausten sijaintia. Perus liikennemäärä- ja keskinopeusaikasarjojen lisäksi syötteinä kokeiltiin kellonaika- ja viikonpäivätietojen antamista eri muodoissa (joko aika- ja päivätietoina tai kyseisten hetkien keskimääräisinä liikennemäärä- ja keskinopeustietoina) sekä nopeuden keskihajontaa. Perusmallit tehtiin ennustamaan liikennetilanne 15 minuuttia eteenpäin. Tarkoitus oli selvittää, kuinka ennusteen laatu muuttuu, kun ennustejakson pituus kasvaa 30 minuuttiin. Perusmallit tehtiin oletuksella, että käytettävissä on tiheästi asennettuja ilmaisimia. Tarkoitus oli kuitenkin selvittää, miten poikkileikkaukset kannattaisi sijoittaa silloin, kun niiden määrä on rajallinen. Alla ovat asiaan liittyvät tutkimushypoteesit.

- Kellonaika- ja päivätiedot parantavat ennusteita.
- Sillä, miten kellonaika- ja päivätiedot annetaan, ei ole merkitystä ennusteen laadun kannalta.
- Nopeuskeskihajontatiedot parantavat etenkin nopeusennusteita.
- Ennusteen tarkkuus huononee ennustejakson pituuden kasvaessa.
- Mitä kauempana lisätietopoikkileikkaus sijaitsee ennustepoikkileikkauksesta, sitä tarkempia ovat ennusteet suhteellisen lyhyistä etäisyyksistä puhuttaessa.
- On olemassa jokin etäisyys, jota lähemmäksi poikkileikkauksia ei kannata sijoittaa silloin, kun niiden määrä on rajallinen.

Selvityksen lopussa neuroverkkomallin antamia ennusteita verrattiin naiivin ennustemallin antamiin tuloksiin. Lisäksi testattiin, kuinka hyvin ruuhka-ajan liikennettä kyettiin ennustamaan. Länsiväylältä saatuja kokemuksia sovellettiin kaksikaistaisessa kohteessa. Alla on asiaan liittyviä tutkimushypoteeseja.

- Neuroverkkomallin antamat ennusteet ovat tarkempia kuin naiivin mallin, koska naiivi malli ei pysty reagoimaan odottamattomiin liikenteenhäiriöihin.
- Ruuhka-ajan havainnoilla opetettu malli antaa parempia ruuhka-ajan ennusteita kuin koko päivän aineistolla opetettu malli.
- Kaksikaistaisella tiellä mallille kannattaa antaa syötteeksi tiedot myös vastaantulevasta liikenteestä.
- Länsiväylältä saatuja kokemuksia voidaan soveltaa suurelta osin suoraan myös kaksikaistaisella tiellä.

6 Optimaalinen neuroverkko

6.1 Perusennustemalli

Länsiväylän aineisto oli käytettävissä sekä poikkileikkaus- että kaistakohtaisina arvoina. Ennustemallin syötetiedoiksi valittiin poikkileikkauskohtainen aineisto, jottei piilokerroksen neuronien määrä pienenesi liikaa. Aineisto jaettiin kolmeen osaan. Yhdestä neljäsosasta tehtiin validointidata, yhdestä neljäsosasta testidata ja puolet datasta jätettiin opetusdataksi. Opetusdatan kooksi saatiin 1 627 havaintoa, validointidatan 813 ja testidatan 814 havaintoa.

Opetusdata koostui niistä havainnoista, joiden perusteella neuroverkko yritti oppia mallinnettavaa ilmiötä. Validointidata taas oli aineisto, jonka perusteella yritettiin havaita, milloin opetus oli edennyt niin pitkälle, ettei neuroverkko enää oppinut yleisiä mallinnettavaan ilmiöön liittyviä ominaisuuksia, vaan alkoi opetella käytetyn opetusjoukon erityispiirteitä. Tämä ilmeni tilanteena, jolloin opetusjoukon virhe pieneni, mutta validointijoukon virhe alkoi kasvaa. Testijoukko oli kolmas aineisto ja sitä ei käytetty opetusprosessissa millään tavalla. Sen tarkoituksena oli testata, kuinka hyvin opetuksessa onnistuttiin eli kuinka hyvin neuroverkko toimi.

Syötedatana kokeiltiin kolmea erilaista poikkileikkauskohtaista aineistoa: ensimmäisessä oli viimeisen 15 minuutin liikennemäärä- ja nopeustiedot kaikilta seitsemältä poikkileikkaukselta (42 syöteparametriä, eli kolme liikennemäärä- ja kolme keskinopeushavaintoa jokaisesta poikkileikkauksesta), toisessakin aineistossa poikkileikkauksia oli seitsemän, mutta aineisto oli ainoastaan viimeiseltä kymmeneltä minuutilta (28 syöteparametriä) ja kolmannessa datassa tiedot olivat jälleen 15 viimeiseltä minuutilta, mutta ainoastaan viideltä poikkileikkaukselta (30 syöteparametriä). Karsituiksi tulivat poikkileikkaukset 14 ja 6. Näin päästiin kokeilemaan vaihtokauppaa vähemmän tiedon, mutta useamman neuronin, ja enemmän tiedon, mutta vähempien neuronien välillä.

Tutkimuksessa aineisto muokattiin kahdella eri tavalla: toinen aineisto skaalattiin välille $[-1, 1]$ ja toinen normeerattiin nollakeskiarvoiseksi ja yksihajontaiseksi. Normeeratulle aineistolle tehtiin lisäksi pääkomponenttianalyysi, jolla sen dimensiota saatiin pienemmäksi. Pääkomponenttianalyysissä karsittiin ne komponentit, jotka selittivät alle prosentin aineiston vaihteluista.

Nyt tehdyssä tutkimuksessa piilokerrosten lukumääräksi valittiin yksi ja se koostui Widrow:n säännön mukaisesta määrästä neuroneja alaspäin pyöristettynä (esim. $N_{hmax} = 4,8 \Rightarrow N_h = 4$). Perusennustemallien syöte- ja piilokerroksen neuronien määrät on esitetty taulukossa 3.

Taulukko 3. Perusennustemallien syöte- ja piilokerroksen neuronien lukumäärät eri datoille. Vastaneuronien määrä on kuusi.

Poikkileikkausten määrä / syöte- datan aikasarjan pituus	Normeerattu data		Skaalattu data	
	Syöteneuronien lkm	Piiloneuronien lkm	Syöteneuronien lkm	Piiloneuronien lkm
7 pl / 15 min	9	9	42	2
7 pl / 10 min	8	10	28	4
5 pl / 15 min	8	10	30	4

Levenberg-Marquardt-algoritmi on osoittautunut nopeimmaksi kohtuullisen kokoisten (korkeintaan muutamia satoja painokertoimia) eteenpäin syöttävien neuroverkkojen opetusmenetelmäksi (Demuth & Beale 1998). Kyseistä algoritmia käytettiin myös tässä tutkimuksessa.

Opetettaessa verkolle annettiin erilaisia lopetuskriteerejä. Näitä olivat opetuskierrosten maksimimäärä, gradientin ja keskineliövirheen minimi sekä validointidatan keskineliövirheen pienenemisen loppuminen. Opetuskierrosten maksimimääräksi asetettiin tuhat, minimigradientiksi 10^{-10} , minimikeskineliövirheeksi 10^{-5} ja validointidatakriteeriksi viisi opetuskierrosta, joiden aikana validointidatan keskineliövirhe oli joko pysynyt samana tai kasvanut. Käytännössä opetus päättyi aina validointidatan keskineliövirheen pienenemisen loppumiseen.

Skaalatun aineiston arvot vaihtelivat välillä [-1, 1]. Tämän takia tulostuskerroksen aktivaatiofunktio voi olla joko lineaarinen (ei vaihteluväliä) tai hyperbolinen tangentti (saa arvot samalta väliltä). Normeerattua aineistoa käyttävän neuroverkon tulostuskerroksen aktivaatiofunktion oli oltava lineaarinen, koska verkon vasteiden arvoa ei ollut rajattu.

Skaalatulle aineistolle kokeiltiin seuraavanlaisia aktivaatiofunktioyhdistelmiä: logistinen – lineaarinen, hyperbolinen tangentti – lineaarinen ja hyperbolinen tangentti – hyperbolinen tangentti. (Funktioista ensimmäinen on piilokerroksen aktivaatiofunktio ja toinen tulostuskerroksen.) Normeeratulle aineistolle kokeiltiin kahta eri yhdistelmää: logistinen – lineaarinen ja hyperbolinen tangentti – lineaarinen.

Parhaan verkkorakenteen löytämiseksi erilaisia aktivaatiofunktioyhdistelmiä testattiin eri aineistoilla. Koska opetusprosessi oli satunnainen (painokertoimien ja harhatermien alkuarvot valittiin satunnaisesti Nguyen ja Widrown menetelmällä), kukin verkko opetettiin kymmenen kertaa ja näistä kymmenestä verkosta valittiin paras. Eri verkkojen keskimääräiset neliövirheet on esitetty taulukossa 4. On huomattava, etteivät normeeratun ja skaalatun aineiston keskineliövirheet ole yhteismitallisia.

Taulukko 4. Keskineliövirheet erilaisille perusennustemalleille. Aineistot ovat joko viideltä tai seitsemältä poikkileikkaukselta (5 pl / 7 pl) 15 tai 10 minuutin ajalta.

	Aktivaatiofunktiot	7 pl / 15 min	7 pl / 10 min	5 pl / 15 min
Normeerattu pääkomponenttianalyysin avulla käsitelty data	Log + Lin	0,156	0,174	0,160
	Tan + Lin	0,154	0,170	0,162
Skaalattu data	Log + Lin	0,00631	0,00630	0,00611
	Tan + Lin	0,00634	0,00603	0,00600
	Tan + Tan	0,00656	0,00628	0,00611

Normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa käyttävillä neuroverkoilla parhaaksi yhdistelmäksi osoittautui verkko, jonka syöteparametrit oli johdettu aineistosta, joka oli viimeiseltä 15 minuutilta seitsemältä poikkileikkaukselta, ja jonka piilokerroksen aktivaatiofunktio oli hyperbolinen tangentti. Piilokerroksen hyperbolinen tangentti yhdistettynä tulostuskerroksen lineaariseen funktioon oli parempi kuin yhdistelmä, jossa piilokerroksen aktivaatiofunktio oli logistinen.

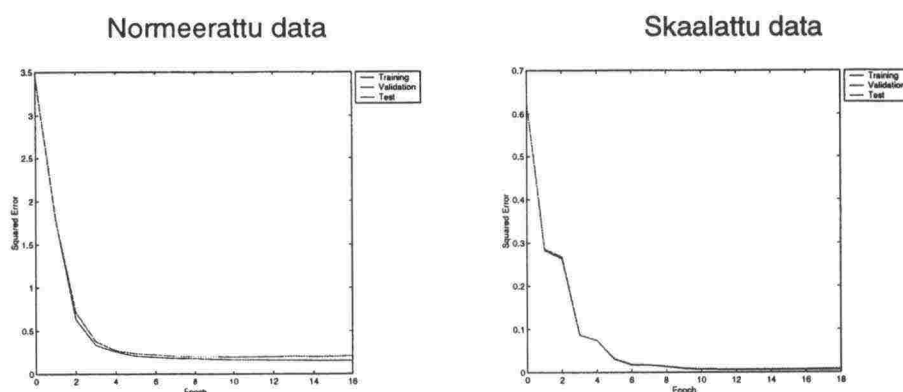
Aktivaatiofunktioyhdistelmästä riippumatta normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa käyttävillä neuroverkoilla aineistoista osoittautui parhaaksi edellä mainittu eniten tietoa sisältävä aineisto. Tiedon vähentäminen piilokerroksen neuronien määrän kasvattamiseksi ei siis kannattanut. Tähän oli varmasti syynä syötetiedon vähentämisen antama pieni suhteellinen lisäys piilokerroksen neuronien määrään (9 vs. 10). Aineistoista toiseksi parhaimpaan tulokseen johti se, joka oli johdettu viimeisen 15 minuutin liikennetiedoista viideltä poikkileikkaukselta. Tämän perusteella vaikuttaa siis siltä, että aikasarjan pituus on tärkeämpi kuin poikkileikkauksien määrä – ainakin kun puhutaan pienistä muutoksista ja poikkileikkaukset ovat lähellä toisiaan.

Skaalattua aineistoa käyttävistä neuroverkoista parhaimmaksi osoittautui verkko, jonka syötedata oli viimeiseltä 15 minuutilta viideltä poikkileikkaukselta, ja jonka piilokerroksen aktivaatiofunktio oli hyperbolinen tangentti ja tulostuskerroksen lineaarinen funktio. Myös skaalatulla aineistolla piilokerroksen aktivaatiofunktiona hyperbolinen tangentti johti parempiin tuloksiin kuin logistinen aktivaatiofunktio yhtä poikkeusta (15 minuutin liikennetiedot seitsemältä poikkileikkaukselta) lukuun ottamatta kaikilla aineistoilla, kun tulostuskerroksen aktivaatiofunktio oli lineaarinen.

Skaalattua aineistoa käyttävissä neuroverkoissa, joissa siis oli huomattavasti enemmän syöteparametrejä (28 – 42 kpl) kuin normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa käyttävissä neuroverkoissa (8 – 9 kpl), voitiin havaita, että neuronien määrää kannatti kasvattaa syötetiedon määrän kustannuksella. Kaikilla aktivaatiofunktioyhdistelmillä poikkileikkauksien määrän tai aikasarjan pituuden pienentäminen kannatti, koska piilokerroksen neuronien määrän kaksinkertaistaminen kahdesta neljään paransi tulosta enemmän kuin tiedon menettäminen huononsi. Myös skaalattua aineistoa käyttävillä neurover-

koilla poikkileikkausten määrän pienentäminen oli aina kannattavampaa kuin aikasarjan lyhentäminen.

Parhaimmiksi osoittautuivat siis neuroverkot, joiden piilokerroksessa oli aktivaatiofunktiona hyperbolinen tangenti ja tulostuskerroksessa lineaarinen funktio. Skaalattua aineistoa käyttävistä neuroverkoista parhaiten menestyi verkko, joka sai syötetiedoikseen viiden poikkileikkauksen liikennetiedot 15 minuutilta, ja normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa käyttävistä neuroverkoista verkko, joka sai syötetiedoikseen seitsemän poikkileikkauksen liikennetiedot 15 minuutilta. Kuvassa 6 on esitetty keskineliövirheen kehitys neuroverkkoja opetettaessa kullekin kolmelle aineistojoukolle. Aineistojen virhetermien käyttäytymisen samanlaisuus viittaa alkuperäisen aineiston onnistuneeseen jakoon opetus-, validointi- ja testidataksi.



Kuva 6. Keskineliövirheen kehitys opetus-, validointi- ja testidatalla neuroverkkoja opetettaessa opetuskierrosten lukumäärän funktiona. Vasemman puoleinen malli perustuu normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon ja oikean puoleinen malli skaalattuun aineistoon.

Taulukossa 6 on esitetty mallien virhetunnuslukuja koko havaintojoukosta laskettuna. Nämä virheet ovat yhteismitallisia ja siten vertailukelpoisia. Tarkasteltaessa keskimääräistä neliövirhettä tai absoluuttisen tai suhteellisen virheen itseisarvoa havaittiin, että skaalattuun aineistoon perustuva malli teki parempia liikennemääräennusteita kuin normeerattuun pääkomponenttianalyysillä käsiteltyyn aineistoon perustuva malli. Nopeusennusteissa tilanne oli päinvastainen - tosin keskimääräisen suhteellisen virheen itseisarvo oli sama molemmille malleille. Jos mallin hyvyttä mitattiin keskimääräisellä absoluuttisella virheellä, havaittiin, että normeerattuun pääkomponenttianalyysillä käsiteltyyn aineistoon perustuva malli oli parempi ensimmäisen viiden minuutin liikennemääräennusteissa, mutta skaalattuun aineistoon perustuva malli seuraavien kymmenen minuutin ennusteissa. Keskimääräistä suhteellista virhettä tarkasteltaessa molemmat mallit toimivat yhtä hyvin (keskimääräinen suhteellinen virhe oli noin nolla prosenttia). Nopeusennusteissa malleilla ei ollut eroja näitä kahta viimeistä virhetermiä tarkasteltaessa.

Taulukko 5. Perusennustemallien virhetermit. Taulukossa normeeratulla aineistolla tarkoitetaan normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa.

Virhetermi	Aikajakso	Liikennemäärä		Nopeus	
		Skaalattu	Normeerattu	Skaalattu	Normeerattu
Keskimääräinen neliövirhe	0 - 5	38 000	46 000	6	5
	5 - 10	45 000	55 000	7	6
	10 - 15	55 000	68 000	7	7
Keskimääräinen absoluuttinen virhe	0 - 5	4	0	0	0
	5 - 10	-1	3	0	0
	10 - 15	2	6	0	0
Keskimääräinen absoluuttisen virheen itseisarvo	0 - 5	150	170	2	1
	5 - 10	160	180	2	1
	10 - 15	180	200	2	2
Keskimääräinen suhteellinen virhe	0 - 5	0 %	0 %	0 %	0 %
	5 - 10	0 %	0 %	0 %	0 %
	10 - 15	0 %	0 %	0 %	0 %
Keskimääräinen suhteellisen virheen itseisarvo	0 - 5	8 %	9 %	2 %	2 %
	5 - 10	9 %	10 %	2 %	2 %
	10 - 15	9 %	11 %	2 %	2 %

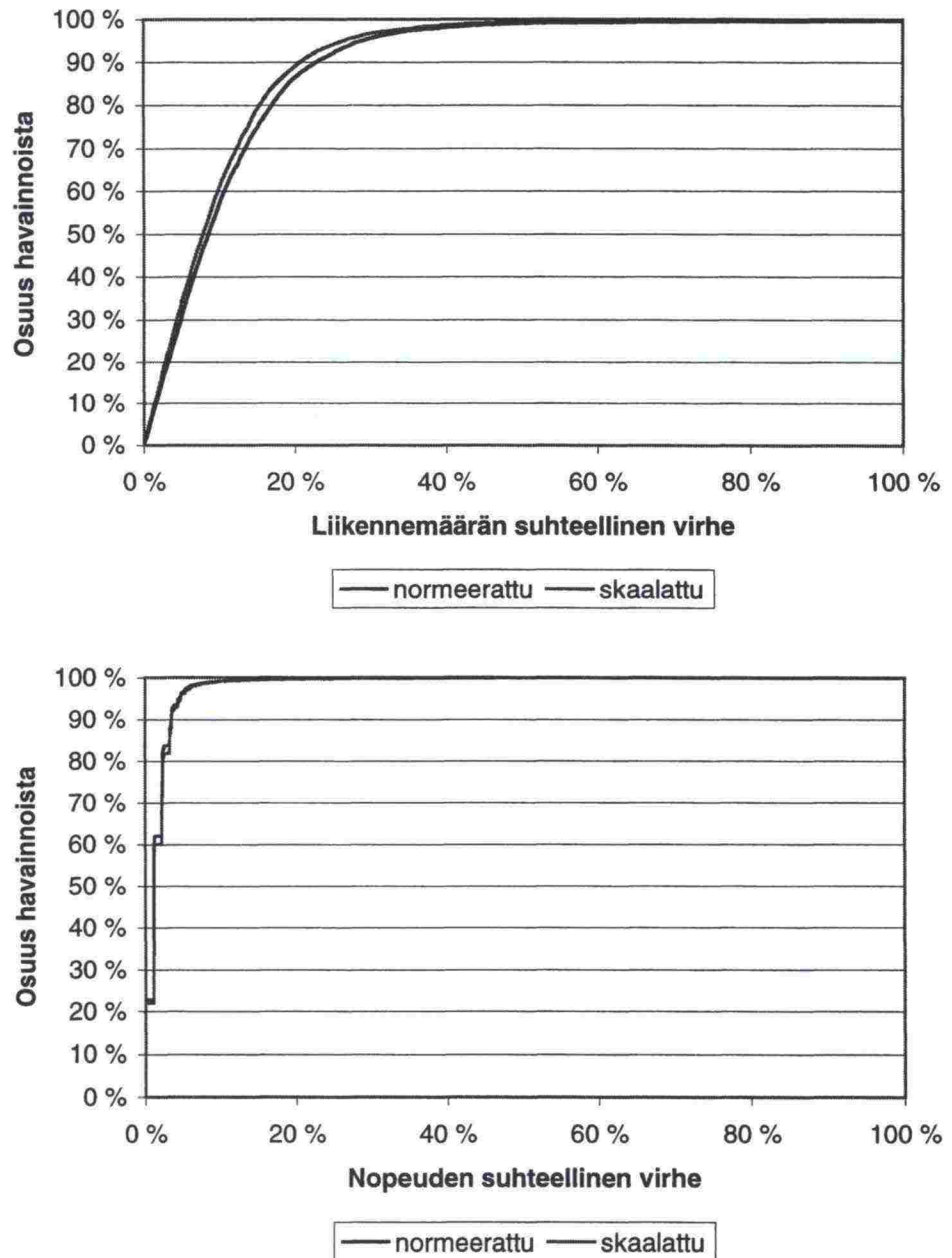
Molempien mallien antamia ennusteita verrattiin tavoitearvoihin myös regressioanalyysin avulla. Analyysissä parametrin ennustetut ja tavoitearvot piirrettiin havaintopisteinä siten, että ennustettu arvo oli y-akselilla ja tavoitearvo x-akselilla. Havaintopisteisiin sovitettiin suora ja ennustettavalle parametrille laskettiin kolme tunnuslukua: suoran kulmakerroin (m), paikka (b), jossa suora leikkaa y-akselin, sekä korrelaatiokerroin (r). Optimaalisessa tilanteessa havaintopisteet asettuvat origon kautta kulkevalle suoralle $y = x$, jolloin $m = 1$, $b = 0$ ja $r = 1$. Saadut tulokset on esitetty taulukossa 6.

Taulukko 6. Perusennustemallin regressioanalyysin tunnusluvut. Tunnusluvuista m tarkoittaa regressiosuoran kulmakerrointa, b pistettä, jossa regressiosuora leikkaa y -akselin, ja r on selitysaste. Taulukossa normeeratulla aineistolla tarkoitetaan normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa.

Aikajakso	Tunnusluku	Liikennemäärä		Nopeus	
		Skaalattu	Normeerattu	Skaalattu	Normeerattu
0 - 5	m	0,95	0,92	0,86	0,76
	b	100	150	13	21
	r	0,98	0,97	0,82	0,85
5 - 10	m	0,94	0,93	0,73	0,69
	b	110	140	24	26
	r	0,97	0,96	0,78	0,81
10 - 15	m	0,92	0,91	0,71	0,71
	b	150	170	25	25
	r	0,96	0,95	0,76	0,77

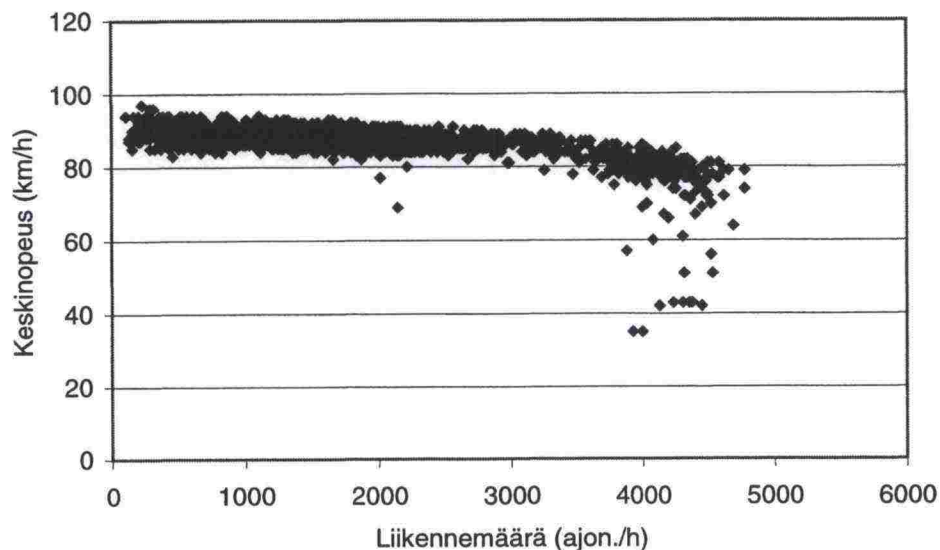
Molemmat mallit ennustivat liikennemääriä (skaalattu data: $R^2 \in [0,93; 0,95]$, normeerattu data: $R^2 \in [0,91; 0,94]$) paremmin kuin nopeuksia (skaalattu data: $R^2 \in [0,58; 0,68]$, normeerattu data: $R^2 \in [0,60; 0,73]$). Skaalattuun aineistoon perustuvan mallin suorat olivat molemmilla muuttujilla lähempänä optimisuoria kuin normeerattuun aineistoon perustuvalla mallilla. Skaalattuun aineistoon perustuvan mallin liikennemääräennusteiden selitysaste oli parempi kuin normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuvan mallin. Nopeusennusteiden selitysasteissa tilanne oli kuitenkin päinvastainen.

Kuvassa 7 on esitetty molemmille malleille suhteellisen virheen itseisarvon summakäyrät sekä liikennemäärä- että nopeusennusteille. Normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuvan mallin liikennemääräennusteista 90 prosentissa suhteellinen virhe oli korkeintaan 23 prosenttia ja skaalattuun aineistoon perustuvalla mallilla 21 prosenttia. Ensiksi mainitulla mallilla 60 prosentissa ja jälkimmäisellä mallilla 63 prosentissa liikennemääräennusteista suhteellinen virhe oli korkeintaan kymmenen prosenttia. Molemmilla malleilla 90 prosentissa nopeusennusteista suhteellinen virhe oli korkeintaan neljä prosenttia ja 99 prosentissa ennusteista virhe oli korkeintaan kymmenen prosenttia.

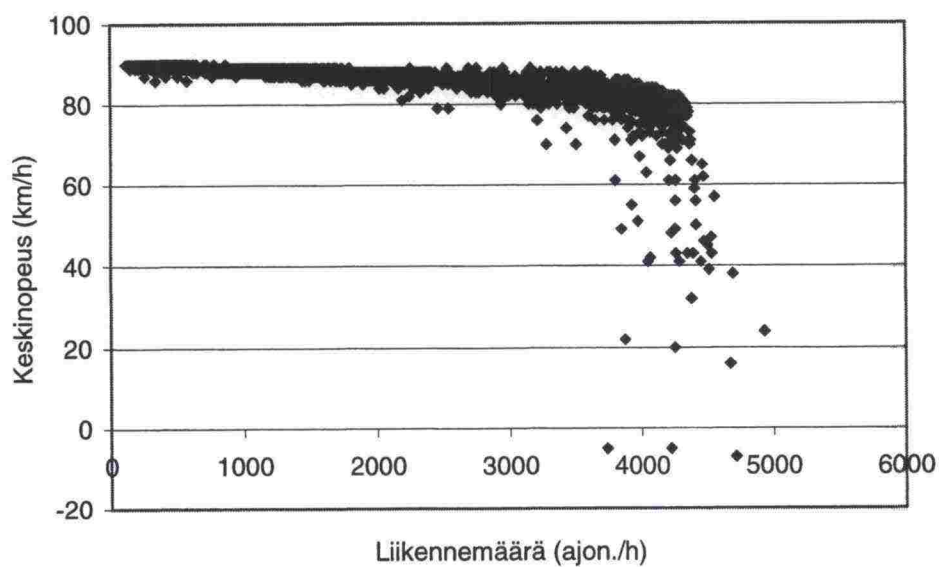


Kuva 7. Liikennemäärän ja keskinopeuden ennusteiden suhteellisten virheiden itseisarvojen summakäyrät skaalattuun ja normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustulle malleille.

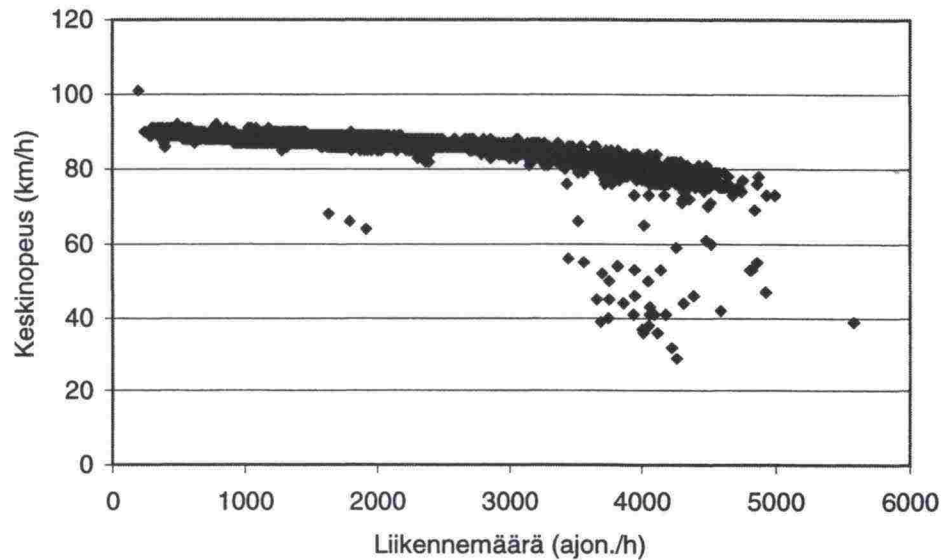
Seuraavissa kuvissa on esitetty alkuperäisestä aineistosta piirretty liikennemäärä-nopeuspisteistö sekä ennustemallien tulosten perusteella piirretyt vastaavat pisteistöt. Kuvista havaitaan, että ennusteiden pisteet osuivat hyvin peruskuvajalle.



Kuva 8. Alkuperäisestä aineistosta piirretty liikennemäärä-keskinopeus-pisteistö.



Kuva 9. Skaalattuun aineistoon perustuvan mallin tekemistä ennusteista piirretty liikennemäärä-keskinopeuspisteistö.



Kuva 10. Normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuvan mallin tekemistä ennusteista piirretty liikennemäärä-keskinopeuspisteistö.

6.2 Yksi malli vs. hajautettu malli

6.2.1 Oma malli kullekin ennustejaksolle

Jos malli jaetaan useampaan osamalliin, joista kukin ennustaa vain osan halutuista parametreista, mallin vastekerroksen neuronien määrä pienenee ja näin ollen neuroverkon piilokerrokseen voidaan saman kokoisella opetusjoukolla laittaa useampia piiloneuroneja kuin neuroverkossa, joka tekee yhdellä kertaa kaikki ennusteet. Muutos on pieni, joten jäi selvittettäväksi, kannattiko osamalleihin jakaminen.

Ensin kokeiltiin mallin jakamista kolmeen neuroverkkoon, joista jokainen teki liikennemäärä- ja keskinopeusennusteen omalle viisiminuuttiselle. Neuroverkoissa käytettiin piilokerroksen aktivaatiofunktiona hyperbolista tangenttia ja tulostuskerroksessa lineaarista funktiota. Kokeiltiin kahta erilaista aineistoa: toisessa oli viimeisen 15 minuutin liikennemäärä- ja nopeustiedot seitsemältä poikkileikkaukselta ja toisessa viideltä poikkileikkaukselta. Aineiston muokkauksessa käytettiin samoja, edellisissä malleissa käytettyjä menetelmiä: skaalausta ja normeeratun aineiston käsittelyä pääkomponenttianalyysin avulla. Neuronien määrät on esitetty taulukossa 7. Vastekerroksen neuronien määrä oli nyt kaksi. On huomattava, että poikkileikkausten määrän vähentämisestä seuraava neuronien määrän lisäys oli nyt suhteellisesti pienempi kuin yhden neuroverkon mallissa.

Taulukko 7. Mallin, jossa kullekin ennustejaksolle on oma osamalli, osamallien syöte- ja piilokerroksen neuronien lukumäärät eri datoille. Kunkin osamallin vastekerroksessa on kaksi neuronia.

Poikkileikkausten lkm / aikasarjan pituus	Normeerattu pääkomponenttianalyysin avulla käsitelty data		Skaalattu data	
	Syötekerroksen neuronien lkm	Piilokerroksen neuronien lkm	Syötekerroksen neuronien lkm	Piilokerroksen neuronien lkm
7 pl / 15 min	9	13	42	3
5 pl / 15 min	8	14	30	4

Taulukossa 8 on esitetty keskineliövirheet erilaisille verkoille. (On huomattava, että normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuvien mallien virheet eivät ole yhteismitallisia skaalattuun aineistoon perustuvien mallien virheiden kanssa.) Ensimmäisellä ennustejaksolla (0 - 5 minuuttia tarkasteluhetkestä eteenpäin) normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa käyttävistä malleista paremmin menestyi neuroverkko, joka käytti lähtötietoja seitsemältä poikkileikkaukselta. Skaalattua aineistoa käyttävistä malleista parhaiten menestyi malli, joka sai syötetietonsa viideltä poikkileikkaukselta.

Taulukko 8. Keskineliövirheet erilaisille malleille, joissa kullekin ennustejaksolle on oma osamalli.

Ennustejakso	0 - 5 minuuttia		5 - 10 minuuttia		10 - 15 minuuttia	
	7 pl / 15 min	5 pl / 15 min	7 pl / 15 min	5 pl / 15 min	7 pl / 15 min	5 pl / 15 min
Normeerattu pääkomponenttianalyysin avulla käsitelty data	0,121	0,124	0,132	-	0,181	-
Skaalattu data	0,00491	0,00485	-	0,00523	-	0,00650

Seuraaville ennustejaksoille mallit tehtiin ensimmäisellä ennustejaksolla parhaisiin tuloksiin johtaneilla aineistoilla. Molemmista malleista ennustejakson 5 - 10 minuuttia tulokset olivat hieman huonompia kuin ennustejakson 0 - 5 minuuttia. Viimeisen ennustejakson (10 - 15 minuuttia tarkasteluhetkestä eteenpäin) tulos oli normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa käyttävässä mallissa selvästi edellisiä huonompia, skaalattua aineistoa käyttävässä mallissa huononeminen oli samaa luokkaa kuin kahden ensimmäisen ennustejakson välillä.

Taulukossa 9 on esitetty mallien tulokset erilaisten virhetermien avulla. Tuloksesta havaitaan, että mallien paremmuus vaihteli parametrien ja virhetermien mukaan. Sen sijaan eri ennustejaksojen väleillä eroja oli vähän. Taulukossa 10 on esitetty regressioanalyysin tunnusluvut. Taulukosta havaitaan, että tulokset olivat hyvin tasaisia. Skaalattua aineistoa käyttävällä mallilla kuitenkin kymmenen tunnuslukua 18:sta oli parempia kuin normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa käyttävällä mallilla, joten skaalattua aineistoa käyttävää mallia voitaneen pitää hieman parempana.

Taulukko 9. Virhetermit malleille, joissa kullekin ennustejaksolle on oma osamalli. Taulukossa normeeratulla aineistolla tarkoitetaan normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa.

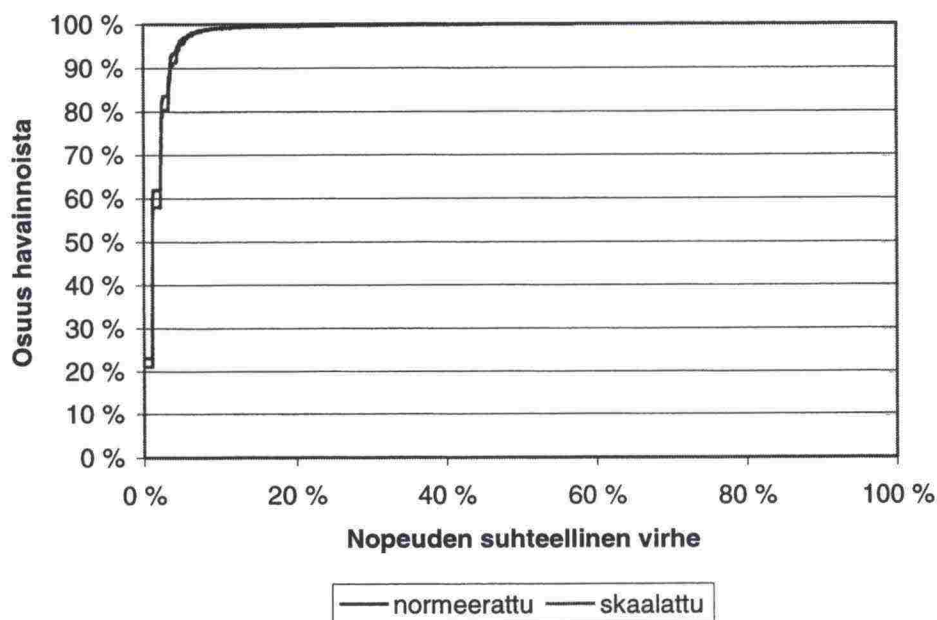
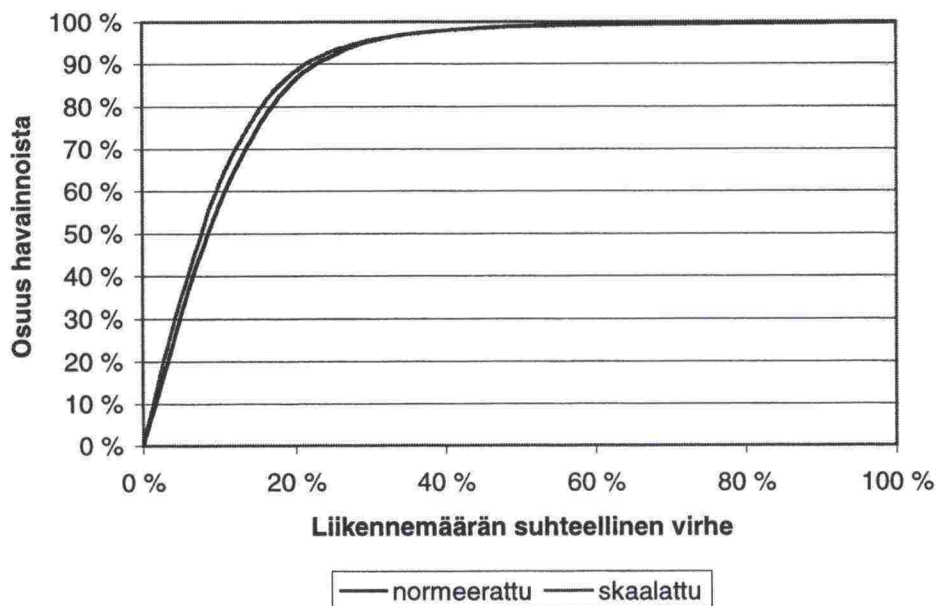
Virhetermi	Aikajakso	Liikennemäärä		Nopeus	
		Skaalattu	Normeerattu	Skaalattu	Normeerattu
Keskimääräinen nelivirhe	0 – 5	38 000	46 000	8	5
	5 – 10	45 000	57 000	7	6
	10 – 15	57 000	71 000	8	8
Keskimääräinen absoluuttinen virhe	0 – 5	20	3	-1	0
	5 – 10	4	19	0	0
	10 – 15	17	-10	0	0
Keskimääräinen absoluuttisen virheen itseisarvo	0 – 5	150	170	2	1
	5 – 10	160	190	2	2
	10 – 15	180	200	2	2
Keskimääräinen suhteellinen virhe	0 – 5	1 %	0 %	-1 %	0 %
	5 – 10	0 %	1 %	0 %	0 %
	10 – 15	1 %	-1 %	0 %	0 %
Keskimääräinen suhteellisen virheen itseisarvo	0 – 5	8 %	9 %	2 %	2 %
	5 – 10	9 %	10 %	2 %	2 %
	10 – 15	10 %	11 %	2 %	2 %

Taulukko 10. Regressioanalyysin tunnusluvut malleille, joissa kullekin ennustejaksolle on oma osamalli. Tunnuslukuista m tarkoittaa regressiosuoran kulmakerrointa, b pistettä, jossa regressiosuora leikkaa y -akselin, ja r on selitysaste. Taulukossa normeeratulla aineistolla tarkoitetaan normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa.

Aikajakso	Tunnusluku	Liikennemäärä		Nopeus	
		Skaalattu	Normeerattu	Skaalattu	Normeerattu
0 – 5	m	0,93	0,94	0,82	0,81
	b	150	110	14	16
	r	0,98	0,97	0,81	0,85
5 – 10	m	0,94	0,96	0,75	0,74
	b	110	100	21	23
	r	0,97	0,96	0,78	0,82
10 – 15	m	0,92	0,90	0,72	0,68
	b	170	190	24	27
	r	0,96	0,95	0,74	0,75

Kuvassa 11 on esitetty mallien, joissa kullekin ennustejaksolle on oma osamalli, liikennemäärä- ja nopeusennusteiden suhteellisten virheiden kumulatiiviset summakäyrät. Normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuvan mallin tekemistä liikennemääräennusteista 90 prosentissa suhteellinen virhe oli korkeintaan 23 prosenttia ja skaalattuun aineistoon perustuvalla mallilla 22 prosenttia. Ensin mainitulla mallilla 59 prosentissa ja jälkimmäisellä mallilla 64 prosentissa liikennemääräennusteista suhteellisen virheen arvo oli korkeintaan kymmenen prosenttia. Molemmilla malleilla 90 prosentissa nopeusennusteista suhteellisen virheen arvo oli korkeintaan neljä prosenttia ja

99 prosentissa ennusteista suhteellisen virheen arvo oli korkeintaan kymmenen prosenttia.



Kuva 11. Liikennemäärän ja keskinopeuden suhteellisten virheiden itseisarvojen kumulatiiviset summakäyrät skaalattuun ja normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuvilla malleilla, joissa kullekin ennustejaksolle on oma osamalli.

6.2.2 Erilliset liikennemäärä- ja nopeusmallit

Toisena versiona kokeiltiin mallin jakamista kahteen neuroverkkoon, joista toinen ennustaa liikennemääriä ja toinen nopeuksia. Neuroverkoissa käytettiin pii-

lokerroksen aktivaatiofunktiona hyperbolista tangenttia ja tulostuskerroksessa lineaarista funktiota. Kokeiltiin kahta erilaista aineistoa: toisessa oli viimeisen 15 minuutin liikennemäärä- ja nopeustiedot seitsemältä poikkileikkaukselta ja toisessa viideltä poikkileikkaukselta. Aineiston muokkauksessa käytettiin samoja, edellisissä malleissa käytettyjä menetelmiä: skaalausta ja normeeratun aineiston käsittelyä pääkomponenttianalyysin avulla. Neuronien määrät on esitetty *taulukossa 11*. Luvut ovat hyvin samanlaiset kuin ennustemallissa, jossa kullekin kolmelle ennustejaksolle oli oma osamallinsa.

Taulukko 11. Mallin, jossa on erilliset osamallit liikennemäärän ja keskinopeuden ennustamiselle, osamallien syöte- ja piilokerroksen neuronien lukumäärät eri aineistoille. Vastekerroksessa on kolme neuronia.

Poikkileikkausten lkm / aikasarjan pituus	Normeerattu pääkomponenttianalyysin avulla käsitelty data		Skaalattu data	
	Syötekerroksen neuronien lkm	Piilokerroksen neuronien lkm	Syötekerroksen neuronien lkm	Piilokerroksen neuronien lkm
7 pl / 15 min	9	12	42	3
5 pl / 15 min	8	13	30	4

Taulukossa 12 on esitetty keskineliövirheet erilaisille verkoille. (On huomattava, että normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuvien mallien virheet eivät ole yhteismitallisia skaalattuun aineistoon perustuvien mallien kanssa.) Liikennemääräennusteissa normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa käyttävistä malleista menestyi parhaimmin neuroverkko, joka käytti lähtötietoja viideltä poikkileikkaukselta. Tehdyn 20:n tällä tavalla käsiteltyä aineistoa käyttävän liikennemäärärien ennustamismallin joukosta viidestä parhaasta mallista tämä oli kuitenkin ainoa viiden poikkileikkauksen malli, kaikki neljä seuraavaa mallia käyttivät seitsemää poikkileikkausta, mikä viittaa siihen, että saatu tulos (paras tulos saadaan viidellä poikkileikkauksella) oli todennäköisesti sattumaa. Näin ollen voitiin sanoa, että normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa käyttävistä neuroverkoista sekä liikennemäärän että nopeuden ennustamisessa parhaiten menestyi seitsemää poikkileikkausta lähtötietoinaan käyttävä neuroverkko. Skaalattua aineistoa käyttävistä malleista sekä liikennemäärä- että nopeusennusteissa parhaiten menestyivät mallit, jotka saivat syötetietonsa seitsemältä poikkileikkaukselta.

Taulukko 12. Keskineliövirheet joko viiden tai seitsemän poikkileikkauksen datoihin perustuville (5 pl / 7 pl) malleille, joissa on erilliset osamallit liikennemäärälle ja keskinopeudelle. Taulukossa normeeratulla aineistolla tarkoitetaan normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa.

Aineisto	Liikennemäärä		Nopeus	
	7 pl / 15 min	5 pl / 15 min	7 pl / 15 min	5 pl / 15 min
Normeerattu	0,0606	0,0602	0,2410	0,2552
Skaalattu	0,00719	0,00725	0,00374	0,00377

Kaikki ennusteet tehtiin siis neuroverkolle, joka sai syötetiedot seitsemältä poikileikkaukselta. *Taulukossa 13* on esitetty mallien virhetermit. Taulukosta havaitaan, että skaalattua aineistoa käyttävä malli teki parempia liikennemääräennusteita kuin normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa käyttävä malli. Tosin keskimääräistä absoluuttista tai suhteellista virhettä tarkastellessa järjestys oli toisinpäin. Myös *taulukossa 14* esitetyistä regressioanalyysin tunnusluvuista havaitaan, että skaalattua aineistoa käyttävän mallin tunnusluvut olivat liikennemääräennusteiden osalta kauttaaltaan hieman parempia kuin normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa käyttävällä mallilla.

Nopeusennusteissa tilanne oli paljon tasaisempi. Suhteellisessa tai keskimääräisessä absoluuttisessa virheessä mallien välillä ei ollut eroa, keskimääräisessä neliövirheessä ja absoluuttisen virheen itseisarvossa normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa käyttävä malli oli hieman parempi kuin skaalattua aineistoa käyttävä malli (*taulukko 13*). *Taulukon 14* tunnuslukujen perusteella skaalattuun aineistoon perustuvan mallin tunnusluvut olivat ensimmäistä ennustejaksoa lukuun ottamatta huonommat kuin normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa käyttävän mallin tunnusluvut.

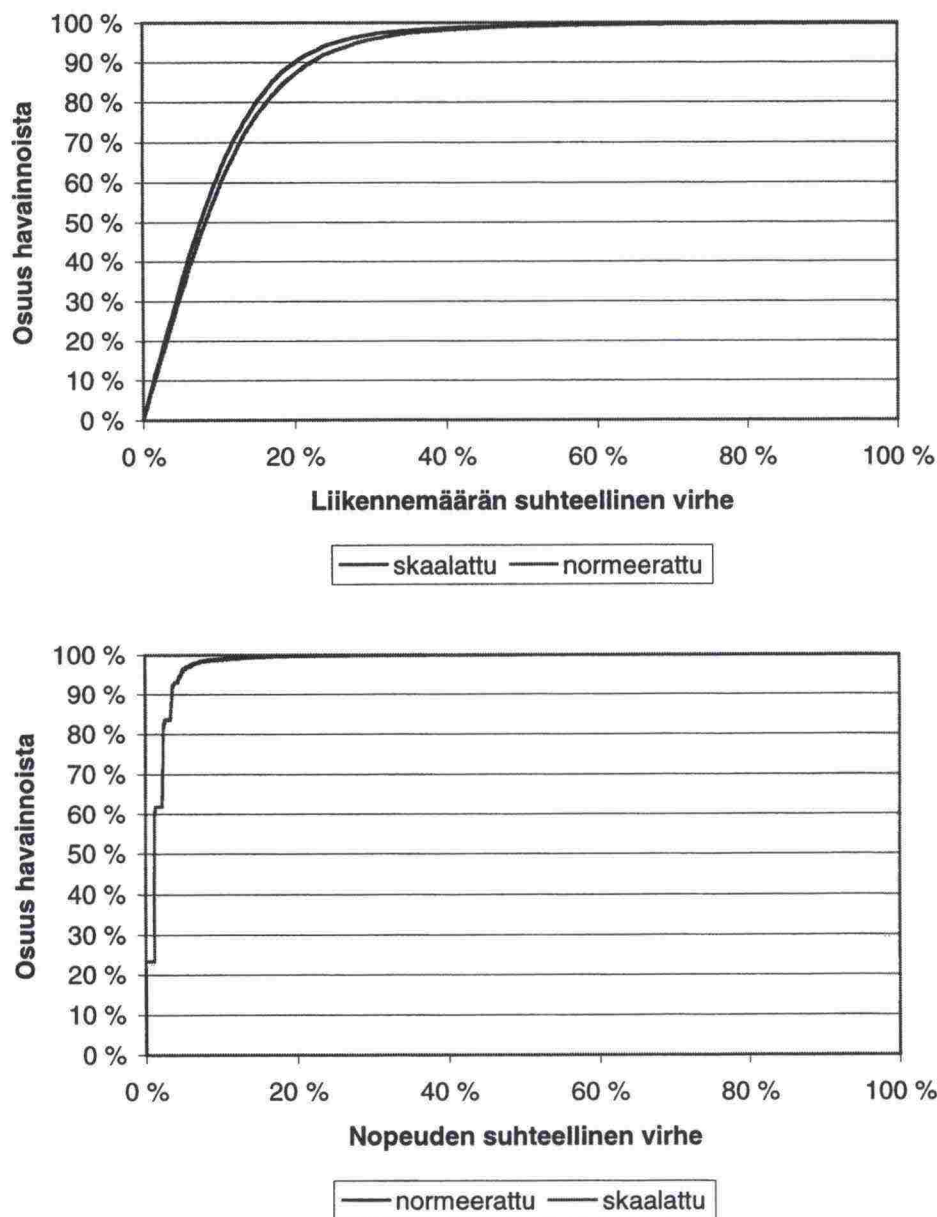
Taulukko 13. Virhetermit ennustemalleille, joissa on erilliset osamallit liikennemäärälle ja keskinopeudelle. Taulukossa normeeratulla aineistolla tarkoitetaan normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa.

Virhetermi	Aikajakso	Liikennemäärä		Nopeus	
		Skaalattu	Normeerattu	Skaalattu	Normeerattu
Keskimääräinen neliövirhe	0 – 5	36 000	43 000	6	5
	5 – 10	44 000	51 000	7	6
	10 – 15	51 000	60 000	7	7
Keskimääräinen absoluuttinen virhe	0 – 5	-11	-7	0	0
	5 – 10	-18	-9	0	0
	10 – 15	-16	-3	0	0
Keskimääräinen absoluuttisen virheen itseisarvo	0 – 5	150	160	1	1
	5 – 10	160	180	2	1
	10 – 15	170	190	2	2
Keskimääräinen suhteellinen virhe	0 – 5	-1 %	0 %	0 %	0 %
	5 – 10	-1 %	0 %	0 %	0 %
	10 – 15	-1 %	0 %	0 %	0 %
Keskimääräinen suhteellisen virheen itseisarvo	0 – 5	8 %	9 %	2 %	2 %
	5 – 10	8 %	9 %	2 %	2 %
	10 – 15	9 %	10 %	2 %	2 %

Taulukko 14. Regressioanalyysin tunnusluvut ennustemalleille, joissa on erilliset osamallit liikennemäärälle ja keskinopeudelle. Tunnusluvuista m tarkoittaa regressiosuoran kulmakerrointa, b pistettä, jossa regressiosuora leikkaa y-akselin, ja r on selitysaste. Taulukossa normeeratulla datalla tarkoitetaan normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa.

Aikajakso	Tunnusluku	Liikennemäärä		Nopeus	
		Skaalattu	Normeerattu	Skaalattu	Normeerattu
0 – 5	m	0,95	0,95	0,85	0,85
	b	91	110	13	13
	r	0,98	0,97	0,83	0,85
5 – 10	m	0,943	0,94	0,710	0,73
	b	90	110	25	23
	r	0,971	0,97	0,778	0,806
10 - 15	m	0,95	0,92	0,68	0,73
	b	110	150	28	24
	r	0,97	0,96	0,77	0,78

Kuvassa 12 on esitetty liikennemäärä- ja nopeusennusteiden suhteellisten virheiden kumulatiiviset summakäyrät ennustemalleille, joissa on erilliset osamallit liikennemäärälle ja keskinopeudelle. Normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuvan mallin tekemistä liikennemääräennusteista 90 prosentissa suhteellinen virhe oli korkeintaan 22 prosenttia ja skaalatun aineistoon perustuvalla mallilla 20 prosenttia. Ensin mainitulla mallilla 61 prosentissa ja jälkimmäisellä mallilla 64 prosentissa liikennemääräennusteista suhteellisen virheen arvo oli korkeintaan kymmenen prosenttia. Molemmilla malleilla 90 prosentissa nopeusennusteista suhteellisen virheen arvo oli korkeintaan neljä prosenttia ja 99 prosentissa ennusteista suhteellisen virheen arvo oli korkeintaan kymmenen prosenttia.



Kuva 12. Liikennemäärän ja keskinopeuden suhteellisten virheiden itseisarvojen summakäyrät skaalattuun ja normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuvilla ennustemalleilla, joissa on erilliset osamallit liikennemäärälle ja keskinopeudelle.

6.2.3 Mallien välinen vertailu

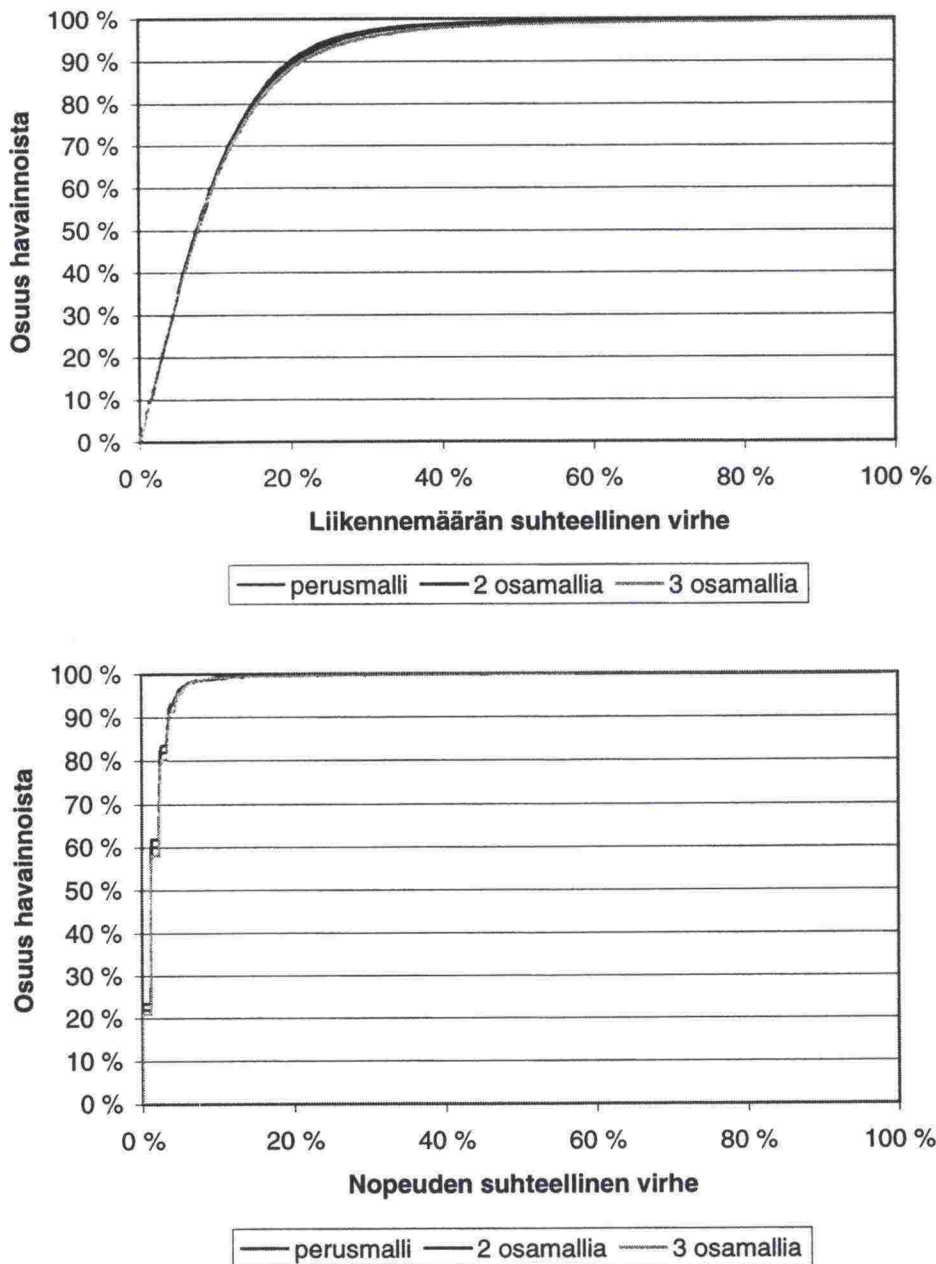
Eräs ratkaistavista ongelmista oli, voidaanko ennustetuloksia parantaa jakamalla malli osamalleihin. Ennustemalli jaettiin kahdella eri tavalla: jakamalla se kahteen (nopeus ja liikennemäärä ennustettiin erikseen) ja kolmeen osamalliin (oma osamalli kullekin viiden minuutin ennustejaksolle). Näitä osamalleihin jaettuja malleja verrattiin jakamattomiin perusmalleihin.

Malleja vertailtiin usean eri virhetermin (keskimääräinen neliövirhe, keskivirhe, keskimääräinen virheen itseisarvo, keskimääräinen suhteellinen virhe ja keskimääräinen suhteellisen virheen itseisarvo) avulla (taulukko 15). Vaikkakin tulokset olivat samaa suuruusluokkaa, malli, jossa kullekin viisiminuuttiselle oli oma osamalli, antoi huonoimmat ennusteet ja malli, jossa nopeus ja liikennemäärä ennustettiin erillisissä osamalleissa, oli paras. Tulos on yllättävä, koska kaikkien osamalleihin jaettujen mallien olisi teoriassa pitänyt olla vähintään yhtä hyviä kuin jakamaton malli. Nyt saatuun tulokseen lienee ollut syynä sattuma.

Taulukko 15. Osamalleihin jaettujen ennustemallien virhetermien vertailu jakamattomaan perusmalliin. Kahteen osamalliin jaetun mallin nopeusennuste perustuu normeerattuun pääkomponenttianalyysiin avulla käsiteltyyn aineistoon, muut skaalattuun aineistoon.

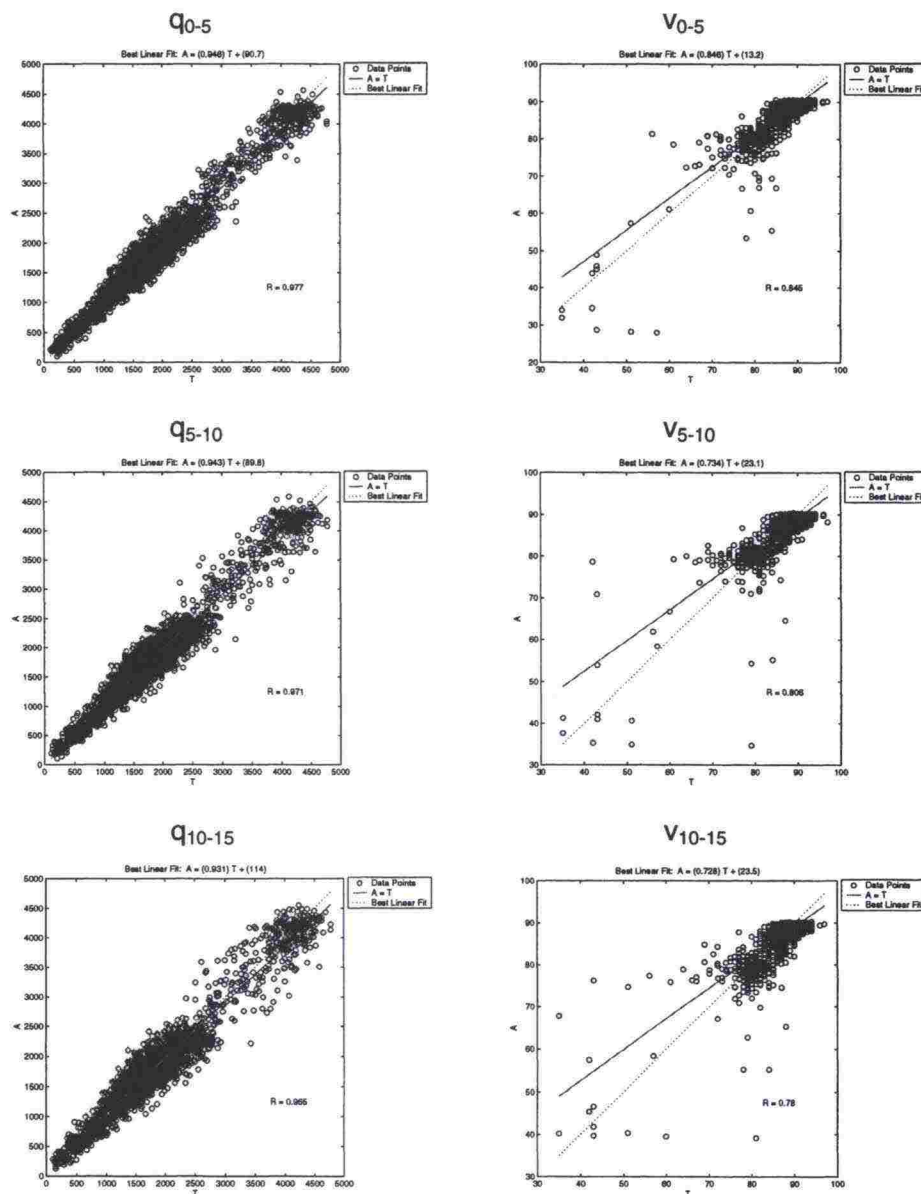
Virhetermi	Aikajakso	Liikennemäärä			Nopeus		
		Perusmalli	2 osamallia	3 osamallia	Perusmalli	2 osamallia	3 osamallia
Keskim. neliövirhe	0 – 5	38 000	36 000	38 000	6	5	8
	5 – 10	45 000	44 000	45 000	7	6	7
	10 – 15	55 000	51 000	57 000	7	7	8
Keskim. absoluuttinen virhe	0 – 5	4	-11	20	0	0	1
	5 – 10	-1	-18	4	0	0	0
	10 – 15	2	-16	17	0	0	0
Keskim. absoluuttisen virheen itseisarvo	0 – 5	150	150	150	2	1	2
	5 – 10	160	160	160	2	1	2
	10 – 15	180	170	180	2	2	2
Keskim. suhteellinen virhe	0 – 5	0 %	-1 %	1 %	0 %	0 %	-1 %
	5 – 10	0 %	-1 %	0 %	0 %	0 %	0 %
	10 – 15	0 %	-1 %	1 %	0 %	0 %	0 %
Keskim. suhteellisen virheen itseisarvo	0 – 5	8 %	8 %	8 %	2 %	2 %	2 %
	5 – 10	9 %	8 %	9 %	2 %	2 %	2 %
	10 – 15	9 %	9 %	10 %	2 %	2 %	2 %

Kuvassa 13 on esitetty liikennemäärän ja keskinopeuden suhteellisen virheen itseisarvon kumulatiiviset summakäyrät eri malleille. Kuvasta nähdään, että vaikka mallit voitiinkin laittaa paremmuusjärjestykseen, mallien väliset erot olivat erittäin pieniä.



Kuva 13. Liikennemäärän ja keskinopeuden suhteellisten virheiden itseisarvojen summakäyrät osamalleihin jakamattomalla perusmallilla (keskimäinen käyrä) sekä kahteen (ylin käyrä) ja kolmeen osamalliin jaetulla mallilla (alin käyrä).

Kuvassa 14 on esitetty regressioanalyysi parhaaksi osoittautuneelle ennustemallille, jonka ensimmäinen osamalli ennusti liikennemäärää skaalatun aineiston avulla ja toinen keskinopeutta normeeratun pääkomponenttianalyysin avulla käsitellyn aineiston avulla.



Kuva 14.

Regressioanalyysi osamalleille, joista toinen ennustaa liikennemäärän ja toinen keskinopeuden. Liikennemäärän ennustamismalli käyttää skaalattua aineistoa ja keskinopeuden ennustamismalli normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa. X-akselilla on "oikea" ja y-akselilla ennustettu arvo. Liikennemäärää on merkitty q:lla ja keskinopeutta v:llä. Alaindeksit viittaavat ennustejaksoon.

7 Muiden tekijöiden optimoiminen

7.1 Päivä- ja kellonaikatiedot lisäsyötteinä

Ennustemallin tuloksia yritettiin parantaa sillä, että syötetietoihin lisättiin päivätyyppi (arkipäivä, lauantai ja sunnuntai/pyhä) ja kellonaika. Päivätyyppi annettiin dummy-muuttujina, joista ensimmäinen sai arvon yksi, kun havainto oli arkipäivältä, toinen arvon yksi, kun havainto oli lauantailta, ja kolmas arvon yksi, kun havainto oli sunnuntailta tai pyhäpäivältä. Muuten muuttujien arvona oli nolla. Kellonaika annettiin desimaalilukuna (esim. klo 12:30 = 12,5).

Opetusdatassa painottuivat ruuhkattoman ajan havainnot, koska ylikysynnän kesto oli suhteellisen lyhyt. Jotta malli pääsi harjoittelemaan "vaikeita" ylikysyntätilanteiden havaintoja useammin kuin ruuhkattomia havaintoja, joita muutenkin oli enemmän, liikenteen aamuhuipun havainnot (klo 7.00 - 9.00) otettiin opetusdataan kahteen kertaan. Tämän datan lisäyksen jälkeen opetusdatan koko oli 3 816 havaintoa.

Taulukossa 16 on vertailtu edellä kuvatun mallin tuloksia perusmallilla saatuihin tuloksiin. Perusmallilla tarkoitetaan tässä yhteydessä mallia, joka oli rakenteeltaan muuten samanlainen kuin edellä kuvattu malli ja joka oli opetettu samalla opetusjoukolla, mutta joka sai syötetiedoikseen ainoastaan liikennemäärä- ja nopeustiedot. Taulukosta havaitaan, että lisätiedot vaikuttivat parantaneen liikennemääräennusteita jonkin verran, mutta keskinopeusennusteisiin ne eivät olleet tuoneet lisää tarkkuutta. (On huomattava, ettei skaalattuun ja normeerattuun aineistoon perustuvia malleja voi verrata tässä toisiinsa, koska virhetermit eivät ole yhteismitallisia.)

Taulukko 16. Keskimääräiset neliövirheet mallille, jossa on käytetty perusyötteen lisäksi lähtötietoina viikonpäivä- ja kellonaikatietoja, ja sen vertailumallille.

	Liikennemäärä		Keskinopeus	
	Vertailu-malli	Lisätieto-malli	Vertailu-malli	Lisätieto-malli
Normeerattu pääkomponenttianalyysin avulla käsitelty data	0,041	0,032	0,22	0,22
Skaalattu data	0,0072	0,0069	0,0047	0,0048

Taulukossa 17 on esitetty skaalattuun ja normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuvien viikonpäivä- ja kellonaikatietoja hyödyntävien mallien virhetermit. Skaalattuun aineistoon perustuva malli antoi kaikkien virhetermien perusteella mitattuna parempia liikennemääräennusteita kuin normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuva malli. Nopeusennusteissa skaalattuun aineistoon perustuva malli antoi pienemmät keskimääräiset neliövirheet, kun taas normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuva malli antoi pienemmät

keskimääräiset absoluuttiset virheet. Muiden virhetermien osalta mallien välillä ei ollut eroa.

Taulukko 17. Viikontäivä- ja kellonaikatietoja hyödyntävien ennustemallien virhetermit. Taulukossa normeeratulla aineistolla tarkoitetaan normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa.

Virhetermi	Aikajakso	Liikennemäärä		Nopeus	
		Skaalattu	Normeerattu	Skaalattu	Normeerattu
Keskimääräinen neliövirhe	0 – 5	1 820 000	1 860 000	33	34
	5 – 10	1 820 000	1 840 000	34	36
	10 – 15	1 820 000	1 850 000	31	33
Keskimääräinen absoluuttinen virhe	0 – 5	80	80	-1	0
	5 – 10	81	82	-1	0
	10 – 15	83	87	-1	0
Keskimääräinen absoluuttisen virheen itseisarvo	0 – 5	1 020	1 030	3	3
	5 – 10	1 020	1 020	3	3
	10 – 15	1 020	1 020	3	3
Keskimääräinen suhteellinen virhe	0 – 5	41 %	41 %	0 %	0 %
	5 – 10	39 %	40 %	0 %	0 %
	10 – 15	38 %	38 %	0 %	0 %
Keskimääräinen suhteellisen virheen itseisarvo	0 – 5	77 %	77 %	4 %	4 %
	5 – 10	76 %	76 %	4 %	4 %
	10 – 15	75 %	75 %	4 %	4 %

Regressioanalyysin tunnusluvut on esitetty taulukossa 18. Tunnusluvuista havaitaan, että normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuva malli antoi parempia liikennemäärä- ja nopeusennusteita kuin skaalattuun aineistoon perustuva malli - suorien kulmakertoimet ja selitysasteet olivat lähempänä yhtä ja regressiosuora leikkasi y-akselin lähempänä origoa.

Taulukko 18. Viikontäivä- ja kellonaikatietoja hyödyntävien ennustemallien regressioanalyysin tunnusluvut. Tunnusluvuista m tarkoittaa regressiosuoran kulmakertoimta, b pistettä, jossa regressiosuora leikkaa y-akselin, ja r on selitysaste. Taulukossa normeeratulla aineistolla tarkoitetaan normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa.

Aikajakso	Tunnusluku	Liikennemäärä		Nopeus	
		Skaalattu	Normeerattu	Skaalattu	Normeerattu
0 – 5	m	0,96	0,97	0,71	0,80
	b	89	66	25	18
	r	0,98	0,98	0,85	0,85
5 – 10	m	0,95	0,97	0,63	0,71
	b	99	65	32	25
	r	0,98	0,98	0,80	0,82
10 – 15	m	0,94	0,96	0,65	0,69
	b	120	90	30	27
	r	0,98	0,98	0,77	0,81

Vertailumallien tekemät virheet on esitetty *taulukossa 19*. Keskimääräisellä neliövirheellä ja absoluuttisen tai suhteellisen virheen itseisarvolla mitattuna normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuva malli antoi parempia liikennemääräennusteita kuin skaalattuun aineistoon perustuva malli, mutta keskimääräisellä absoluuttisella tai suhteellisella virheellä mitattuna tilanne oli päinvastainen. Skaalattuun aineistoon perustuva malli teki parempia nopeusennusteita kuin normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuva malli, mutta ero oli todella pieni.

Taulukko 19. Päivä- ja kellonaikatiedot saavan mallin vertailumallien virhetermit. Taulukossa normeeratulla aineistolla tarkoitetaan normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa.

Virhetermi	Aikajakso	Liikennemäärä		Nopeus	
		Skaalattu	Normeerattu	Skaalattu	Normeerattu
Keskimääräinen neliövirhe	0 – 5	1 850 000	1 820 000	31	34
	5 – 10	1 850 000	1 820 000	36	36
	10 – 15	1 840 000	1 820 000	34	33
Keskimääräinen absoluuttinen virhe	0 – 5	82	84	0	0
	5 – 10	79	85	0	0
	10 – 15	83	92	0	0
Keskimääräinen absoluuttisen virheen itseisarvo	0 – 5	1 030	1 010	3	3
	5 – 10	1 020	1 010	3	3
	10 – 15	1 020	1 010	3	3
Keskimääräinen suhteellinen virhe	0 – 5	41 %	41 %	0 %	0 %
	5 – 10	39 %	40 %	0 %	0 %
	10 – 15	39 %	39 %	0 %	0 %
Keskimääräinen suhteellisen virheen itseisarvo	0 – 5	78 %	77 %	4 %	4 %
	5 – 10	76 %	76 %	4 %	4 %
	10 – 15	75 %	75 %	4 %	4 %

Vertailumallien regressioanalyysin tulokset on esitetty *taulukossa 20*. Taulukosta havaitaan, että skaalattuun aineistoon perustuvan mallin tunnusluvut olivat liikennemääräennusteissa kauttaaltaan parempia kuin normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuvalla mallilla. Nopeusennusteissa tilanne oli päinvastainen.

Taulukko 20. Päivä- ja kellonaikatiedot saavan mallin vertailumallien regressioanalyysin tunnusluvut. Tunnusluvuista m tarkoittaa regressiosuoran kulmakerrointa, b pistettä, jossa regressiosuora leikkaa y-akselin, ja r on selitysaste. Ihannetapauksessa $m = r = 1$ ja $b = 0$. Taulukossa normeeratulla aineistolla tarkoitetaan normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa.

Aikajakso	Tunnusluku	Liikennemäärä		Nopeus	
		Skaalattu	Normeerattu	Skaalattu	Normeerattu
0 - 5	m	0,97	0,95	0,72	0,80
	b	71	100	25	17
	r	0,98	0,98	0,82	0,86
5 - 10	m	0,96	0,95	0,65	0,70
	b	77	110	30	26
	r	0,98	0,98	0,79	0,84
10 - 15	m	0,95	0,94	0,70	0,70
	b	100	140	26	26
	r	0,98	0,97	0,78	0,81

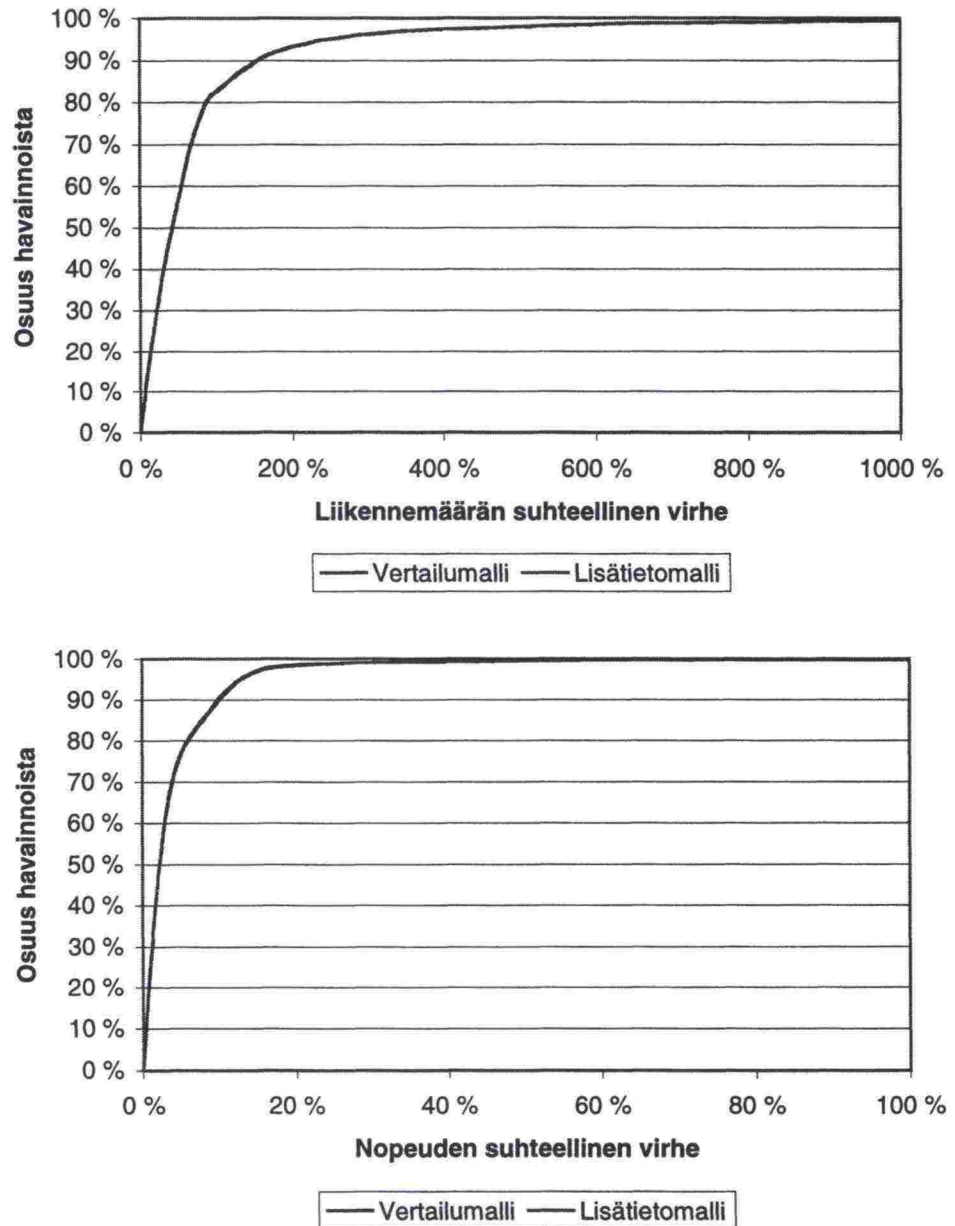
Mallit olivat käytännössä samanarvoisia ja kahdesta vaihtoehdosta valittiin se, jonka tekemät ennusteet olivat lähempänä mitattuja arvoja. Jatkossa siis oletetaan, että ominaisuutta, että ennustettu arvo on lähellä todellista, pidetään arvokkaampana kuin sitä, ettei mallilla ole taipumusta yli- tai aliarvioida ennusteita. Tällä perusteella sekä viikonpäivä- ja kellonaikatietoja hyväksikäyttävistä malleista että niiden vertailumalleista vertailuun valittiin liikennemääräennusteiden osalta skaalattuun aineistoon perustuvat mallit ja keskinopeusennusteiden osalta normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuvat mallit (taulukot 17 - 20).

Malleille laskettiin kunkin virhetermin keskiarvo, joita verrattiin keskenään (taulukko 21). Mallien paremmuus vaihteli liikennemääräennusteissa riippuen siitä, mitä virhetermiä tarkastellaan. Nopeusennusteissa vertailumallin virheet olivat pienemmät kuin lisätietomallin. Mallien väliset erot olivat absoluuttisesti hyvin pieniä, samoin kuin suurelta osin myös suhteellisesti. Osa eroista oli prosenteissa huomattavia, vaikkei eroissa suuria absoluuttisia eroja ollutkaan.

Taulukko 21. Skaalattuun aineistoon perustuvien viikompäivä- ja kellonaikatietoja hyväksikäyttävien mallien sekä niiden normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuvan liikennemäärän vertailuennustemallin ja skaalattuun aineistoon perustuvan keskinopeuden vertailuennustemallin keskimääräiset virhetermit. Lisätieto- ja vertailumallin tekemien virheiden välinen ero on positiivinen, kun lisätietomalli on parempi kuin vertailumalli.

Virhetermi	Liikennemäärä			Nopeus		
	Lisätieto-malli	Ero	Vertailu-malli	Lisätieto-malli	Ero	Vertailu-malli
Keskimääräinen neliövirhe	1 820 000	0 %	1 850 000	34	-1 %	34
Keskimääräinen absoluuttinen virhe	82	-1 %	81	0	-14 %	0
Keskimääräinen absoluuttisen virheen itseisarvo	1 020	0 %	1 020	3	-1 %	3
Keskimääräinen suhteellinen virhe	40 %	0 %	40 %	0 %	-60 %	0 %
Keskimääräinen suhteellisen virheen itseisarvo	76 %	1 %	76 %	4 %	-1 %	4 %

Kuvassa 15 on esitetty viikompäivä- ja kellonaikatietoja hyväksikäyttävien mallien sekä niiden vertailumallien skaalattuun aineistoon perustuvien liikennemääräennusteiden sekä normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuvien keskinopeusennusteiden suhteellisten virheiden summakäyrät. 90 prosentissa molempien ennustamismallien liikennemääräennusteista suhteellinen virhe oli korkeintaan 149 prosenttia. Molempien mallien ennusteista 17 prosentissa suhteellinen virhe oli korkeintaan kymmenen prosenttia. Molempien mallien nopeusennusteista 91 prosentissa suhteellinen virhe oli korkeintaan kymmenen prosenttia.



Kuva 15. Liikennemäärän ja keskinopeuden suhteellisten virheiden it-seisarvojen summakäyrät skaalattuun aineistoon perustuvilla viikonpäivä- ja kellonaikatietoja hyväksikäyttävillä malleilla (lisätietomallit) sekä niiden normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuvalle liikennemäärän vertailuennustemallille ja skaalattuun aineistoon perustuvalle keskinopeuden vertailuennustemallille.

Toinen vaihtoehto päivä- ja kellonaikatietojen hyväksikäyttöön oli antaa mallille syötteeksi keskimääräiset kyseisenä päivänä (arkipäivä, lauantai tai sunnuntai) kysytyllä hetkellä ennustepoikkileikkauksessa havaitut liikennemäärä- ja nopeusestimaatit (liikennemäärien ennustemallille kolme liikennemääräestimaattia ja nopeusmallille kolme nopeusestimaattia). Myös näille malleille tehtiin vertailu-

mallit, jotka olivat rakenteeltaan samanlaisia ja jotka perustuivat samaan ope-
tusjoukkoon - tosin ilman lisätietoja.

Liikennemäärä- ja nopeusestimaatteja hyödyntävien mallien virhetermit (*taulukko 22*) olivat samaa suuruusluokkaa päivä- ja kellonaikatietoja suoraan käyttävi-
en mallien (*taulukko 17*) kanssa. Keskimääräisellä neliövirheellä tai keskimää-
räisellä absoluuttisen tai suhteellisen virheen itseisarvolla tarkasteltuna normee-
rattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuvat liiken-
nemääräennusteet olivat parempia kuin skaalattuun aineistoon perustuvat en-
nusteet. Keskimääräisen absoluuttisen tai suhteellisen virheen itseisarvolla tar-
kasteltuna tilanne oli päinvastainen. Normeerattuun pääkomponenttianalyysin
avulla käsiteltyyn aineistoon perustuva malli antoi parempia nopeusennusteita
kuin skaalattuun aineistoon perustuva malli.

*Taulukko 22. Liikennemäärä- ja nopeusestimaatteja hyödyntävien ennuste-
mallien virhetermit. Taulukossa normeeratulla aineistolla tarkoi-
tetaan normeerattua pääkomponenttianalyysin avulla käsiteltyä
aineistoa.*

Virhetermi	Aikajakso	Liikennemäärä		Nopeus	
		Skaalattu	Normeerattu	Skaalattu	Normeerattu
Keskimääräinen neliövirhe	0 – 5	1 860 000	1 850 000	35	34
	5 – 10	1 860 000	1 850 000	39	35
	10 – 15	1 850 000	1 840 000	37	33
Keskimääräinen absoluuttinen virhe	0 – 5	80	82	-1	0
	5 – 10	80	84	-1	0
	10 – 15	84	89	-1	0
Keskimääräinen absoluuttisen vir- heen itseisarvo	0 – 5	1 030	1 020	3	3
	5 – 10	1 020	1 020	3	3
	10 – 15	1 020	1 020	3	3
Keskimääräinen suhteellinen virhe	0 – 5	41 %	41 %	0 %	0 %
	5 – 10	40 %	40 %	-1 %	0 %
	10 – 15	38 %	39 %	-1 %	0 %
Keskimääräinen suhteellisen virheen itseisarvo	0 – 5	78 %	77 %	4 %	4 %
	5 – 10	76 %	76 %	4 %	4 %
	10 – 15	75 %	75 %	4 %	4 %

Taulukossa 23 on esitetty liikennemäärä- ja nopeusestimaatteja hyödyntävien
ennustemallien regressioanalyysin tunnusluvut. Skaalattuun aineistoon perustu-
van liikennemääriä ennustavan mallin tunnusluvut olivat parempia kuin normee-
rattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuvan mal-
lin. Nopeusennusteiden osalta tilanne ei ollut yhtä selvä. Kahden ensimmäisen
ennustejakson tunnusluvut olivat normeerattuun pääkomponenttianalyysin
avulla käsiteltyyn aineistoon perustuvalla mallilla paremmat, mutta viimeisessä
ennustejaksossa tilanne oli päinvastainen.

Taulukko 23. Liikennemäärä- ja nopeusestimaatteja hyödyntävien ennustemallien regressioanalyysin tunnusluvut. Tunnusluvuista m tarkoittaa regressiosuoran kulmakerrointa, b pistettä, jossa regressiosuora leikkaa y-akselin, ja r on selitysaste. Taulukossa normeeratulla aineistolla tarkoitetaan normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa.

Aikajakso	Tunnusluku	Liikennemäärä		Nopeus	
		Skaalattu	Normeerattu	Skaalattu	Normeerattu
0 – 5	m	0,97	0,96	0,95	0,81
	b	57	75	4	16
	r	0,99	0,98	0,86	0,86
5 – 10	m	0,97	0,96	0,71	0,70
	b	56	80	25	26
	r	0,99	0,98	0,82	0,84
10 – 15	m	0,97	0,95	0,56	0,69
	b	63	100	38	27
	r	0,99	0,98	0,81	0,81

Vertailumallien virhetermit on esitetty taulukossa 24. Taulukosta havaitaan, että skaalattuun aineistoon perustuva malli teki paremmat liikennemääräennusteet kuin normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuva malli. Nopeusennusteissa tilanne oli päinvastainen.

Taulukko 24. Liikennemäärä- ja nopeusestimaatteja hyödyntävien ennustemallien vertailumallien virhetermit. Taulukossa normeeratulla aineistolla tarkoitetaan normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa.

Virhetermi	Aikajakso	Liikennemäärä		Nopeus	
		Skaalattu	Normeerattu	Skaalattu	Normeerattu
Keskimääräinen neliövirhe	0 – 5	1 860 000	1 840 000	44	33
	5 – 10	1 840 000	1 820 000	43	33
	10 – 15	1 840 000	1 820 000	36	32
Keskimääräinen absoluuttinen virhe	0 – 5	90	80	0	0
	5 – 10	88	78	0	0
	10 – 15	90	84	0	0
Keskimääräinen absoluuttisen virheen itseisarvo	0 – 5	1 030	1 020	3	3
	5 – 10	1 020	1 010	3	3
	10 – 15	1 020	1 010	3	3
Keskimääräinen suhteellinen virhe	0 – 5	41 %	41 %	0 %	0 %
	5 – 10	40 %	39 %	0 %	0 %
	10 – 15	39 %	38 %	0 %	0 %
Keskimääräinen suhteellisen virheen itseisarvo	0 – 5	78 %	77 %	4 %	4 %
	5 – 10	76 %	75 %	4 %	4 %
	10 – 15	75 %	74 %	4 %	4 %

Vertailumallien regressioanalyysin tulokset on esitetty taulukossa 25. Taulukosta havaitaan, että skaalattuun aineistoon perustuvan liikennemäärän ennustamismallin tunnusluvut olivat parempia kuin normeerattuun pääkomponenttianalyysin

avulla käsiteltyyn aineistoon perustuvan mallin. Nopeusennusteiden osalta tilanne oli samansuuntainen kahden ensimmäisen ennustejakson osalta selityssastetta lukuun ottamatta. Viimeisessä ennustejaksossa normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuvan mallin tunnusluvut olivat parempia kuin skaalattuun aineistoon perustuvan mallin.

Taulukko 25. Liikennemäärä- ja nopeusestimaatteja hyödyntävien ennustemallien vertailumallien regressioanalyysin tunnusluvut. Tunnusluvuista m tarkoittaa regressiosuoran kulmakerrointa, b pistettä, jossa regressiosuora leikkaa y -akselin, ja r on selitysaste. Ihanetapauksessa $m = r = 1$ ja $b = 0$. Taulukossa normeeratulla datalla tarkoitetaan normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa.

		Liikennemäärä		Nopeus	
Aikajakso	Tunnusluku	Skaalattu	Normeerattu	Skaalattu	Normeerattu
0 – 5	m	0,97	0,95	0,98	0,77
	b	82	99	1	20
	r	0,98	0,98	0,75	0,85
5 – 10	m	0,96	0,95	0,67	0,65
	b	86	94	29	30
	r	0,98	0,98	0,76	0,82
10 - 15	m	0,95	0,94	0,60	0,65
	b	110	120	35	31
	r	0,98	0,97	0,76	0,79

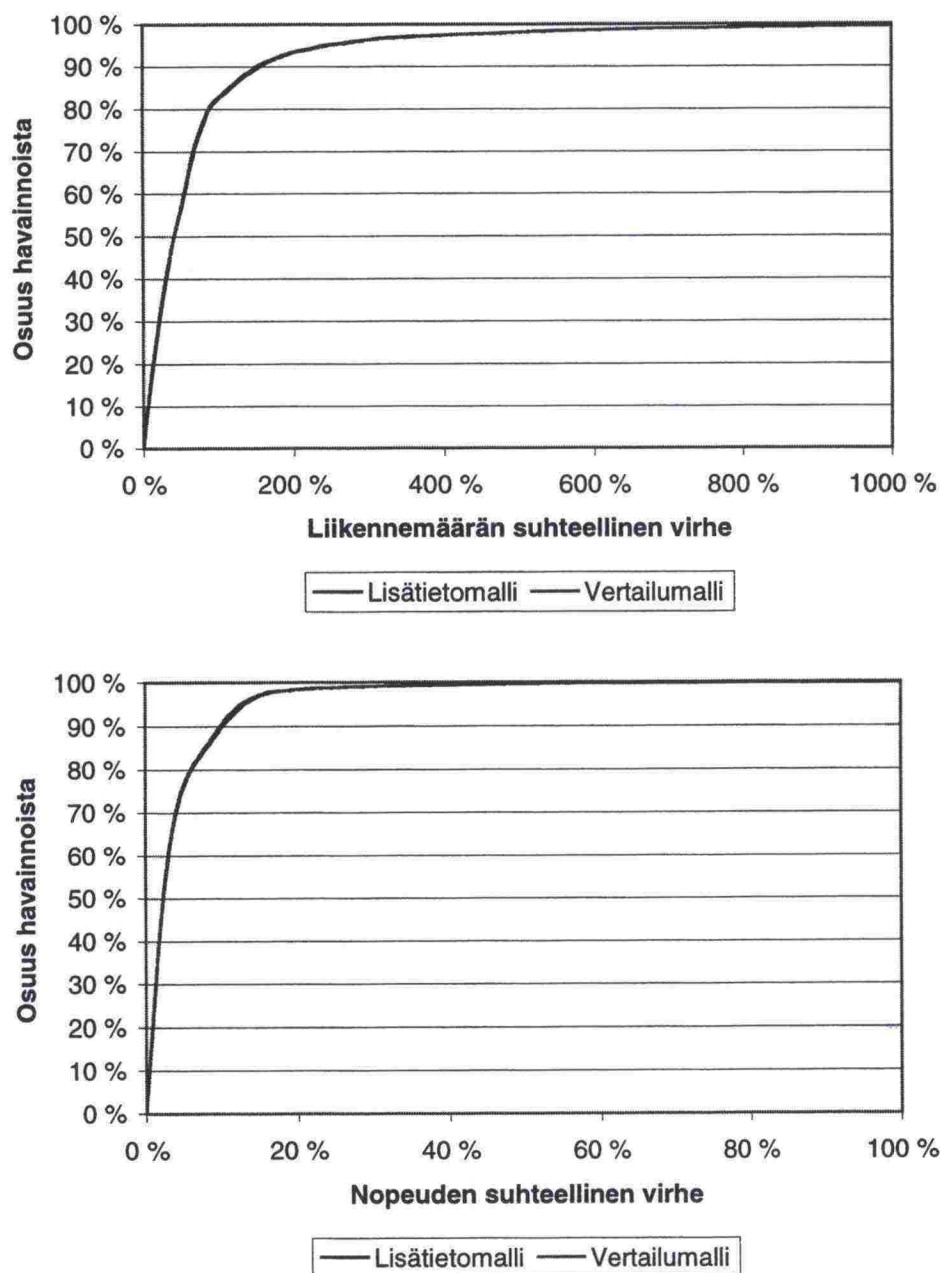
Koska ominaisuutta, että ennustettu arvo on lähellä todellista, päätettiin pitää arvokkaampana kuin sitä, ettei mallilla ole taipumusta yli- tai aliarvioida ennusteita, liikennemäärä- ja nopeusestimaatteja hyväksikäyttävistä malleista valittiin vertailuun liikennemääräennusteiden osalta skaalattuun dataan perustuva malli ja keskinopeusennusteiden osalta normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuva malli (taulukot 22 - 23). Vertailumalleista valittiin molempien ennusteiden osalta normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuvat mallit (taulukot 24 - 25).

Malleille laskettiin kunkin virhetermin keskiarvo, joita verrattiin keskenään (taulukko 26). Vertailumallit osoittautuivat tällä tavalla tarkasteltuna paremmiksi kuin lisätietomallit. Mallien väliset erot olivat suurelta osin suhteellisen pieniä. Osa nopeusennusteiden eroista oli prosenteissa huomattavia, vaikkei eroissa suuria absoluuttisia eroja ollutkaan.

Taulukko 26. Normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuvien liikennemäärä- ja nopeusestimaatteja hyväksikäyttävien mallien sekä niiden normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuvan liikennemäärän vertailuennustemallin ja skaalattuun aineistoon perustuvan keskinopeuden vertailuennustemallin keskimääräiset virhetermit.

Virhetermi	Liikennemäärä			Nopeus		
	Lisätieto- malli	Ero	Vertailu- malli	Lisätieto- malli	Ero	Vertai- lumalli
Keskimääräinen neliövirhe	1 860 000	-2 %	1 830 000	34	-5 %	33
Keskimääräinen absoluuttinen virhe	81	-1 %	80	0	-32 %	0
Keskimääräinen absoluuttisen virheen itseisarvo	1 020	-1 %	1 010	3	-2 %	3
Keskimääräinen suhteellinen virhe	40 %	0 %	39 %	0 %	-169 %	0 %
Keskimääräinen suhteellisen virheen itseisarvo	76 %	-1 %	76 %	4 %	-1 %	4 %

Kuvassa 16 on esitetty liikennemäärä- ja nopeusestimaatteja hyväksikäyttävien mallien (lisätietomallit) skaalattuun aineistoon perustuvien liikennemääräennusteiden sekä normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuvien nopeusennusteiden sekä niiden vertailumallien normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuvien liikennemäärä- ja nopeusennusteiden suhteellisten virheiden kumulatiiviset summakäyrät. 90 prosentissa lisätietomallin liikennemääräennusteista suhteellinen virhe oli korkeintaan 151 prosenttia ja vertailumallin liikennemääräennusteista 148 prosenttia. Lisätietomallin liikennemääräennusteista 16 prosentissa suhteellinen virhe oli korkeintaan kymmenen prosenttia. Vertailumallilla vastaava osuus oli 17 prosenttia. Lisätietomallin nopeusennusteista 91 prosentissa suhteellinen virhe oli korkeintaan kymmenen prosenttia ja vertailumallilla vastaava osuus oli 92 prosenttia.



Kuva 16.

Liikennemäärän ja keskinopeuden suhteellisten virheiden itseisarvojen summakäyrät normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuvilla liikennemäärä- ja nopeusestimaatteja hyväksikäyttävillä malleilla (lisätietomallit) sekä niiden normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuvalla liikennemäärän vertailuennustemallilla ja skaalattuun aineistoon perustuvalla keskinopeuden vertailuennustemallilla.

7.2 Nopeuskeskihajonta lisäsyötteenä

Päivä- ja kellonaikatiedot eivät parantaneet nopeusennusteita. Nopeuden keskihajonta muuttuu nopeuden muutoskohdissa (Innamaa ym. 2000) ja tästä syystä

kokeiltiin, parantaako nopeuskeskihajontatieto ennusteita. Malleille annettiin edellisen 15 minuutin seitsemän poikkileikkauksen liikennemäärä- ja nopeustietojen lisäksi ennustepoikkileikkauksen nopeuskeskihajontatiedot. Malleille tehtiin vertailumallit, jotka olivat rakenteeltaan samanlaisia ja jotka perustuivat samaan opetusjoukkoon - tosin ilman keskihajontatietoja.

Nopeuskeskihajontatietoja hyödyntävien mallien virhetermit on esitetty taulukossa 22. Skaalattuun aineistoon perustuva malli teki parempia liikennemääräennusteita kuin normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuva malli. Nopeusennusteissa tilanne oli päinvastainen.

Taulukko 27. Nopeuskeskihajontatietoja hyödyntävien ennustemallien virhetermit. Taulukossa normeeratulla aineistolla tarkoitetaan normeerattua pääkomponenttianalyysin avulla käsiteltä aineistoa.

Virhetermi	Aikajakso	Liikennemäärä		Nopeus	
		Skaalattu	Normeerattu	Skaalattu	Normeerattu
Keskimääräinen neliövirhe	0 – 5	35 000	44 000	5	5
	5 – 10	43 000	53 000	6	6
	10 – 15	51 000	62 000	7	7
Keskimääräinen absoluuttinen virhe	0 – 5	9	6	0	0
	5 – 10	3	6	0	0
	10 – 15	6	4	-1	0
Keskimääräinen absoluuttisen virheen itseisarvo	0 – 5	150	160	2	1
	5 – 10	160	180	2	2
	10 – 15	170	190	2	2
Keskimääräinen suhteellinen virhe	0 – 5	2 %	2 %	0 %	0 %
	5 – 10	2 %	3 %	0 %	0 %
	10 – 15	2 %	2 %	0 %	0 %
Keskimääräinen suhteellisen virheen itseisarvo	0 – 5	9 %	10 %	2 %	2 %
	5 – 10	10 %	11 %	2 %	2 %
	10 – 15	11 %	12 %	2 %	2 %

Vertailumallien virhetermit on esitetty taulukossa 24. Taulukosta havaitaan, että skaalattuun aineistoon perustuva malli teki parempia liikennemääräennusteita kuin normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuva malli. Nopeusennusteissa skaalattuun aineistoon perustuva malli oli keskimääräisellä neliövirheellä mitattuna hieman parempi kuin normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuva malli, mutta keskimääräisellä absoluuttisella virheellä mitattuna tilanne oli päinvastainen. Muissa virhetermeissä mallien välillä ei ollut eroa.

Taulukko 28. Nopeuskeskihajontatietoja hyödyntävien ennustemallien vertailumallien virhetermit. Taulukossa normeeratulla aineistolla tarkoitetaan normeerattua pääkomponenttianalyysin avulla käsitellyä aineistoa.

Virhetermi	Aikajakso	Liikennemäärä		Nopeus	
		Skaalattu	Normeerattu	Skaalattu	Normeerattu
Keskimääräinen neliövirhe	0 – 5	36 000	43 000	5	5
	5 – 10	43 000	51 000	6	6
	10 – 15	52 000	60 000	6	7
Keskimääräinen absoluuttinen virhe	0 – 5	1	8	0	0
	5 – 10	-3	5	0	0
	10 – 15	3	11	-1	0
Keskimääräinen absoluuttisen virheen itseisarvo	0 – 5	150	160	2	2
	5 – 10	160	180	2	2
	10 – 15	180	190	2	2
Keskimääräinen suhteellinen virhe	0 – 5	2 %	2 %	0 %	0 %
	5 – 10	2 %	2 %	0 %	0 %
	10 – 15	2 %	3 %	0 %	0 %
Keskimääräinen suhteellisen virheen itseisarvo	0 – 5	9 %	10 %	2 %	2 %
	5 – 10	10 %	11 %	2 %	2 %
	10 – 15	11 %	12 %	2 %	2 %

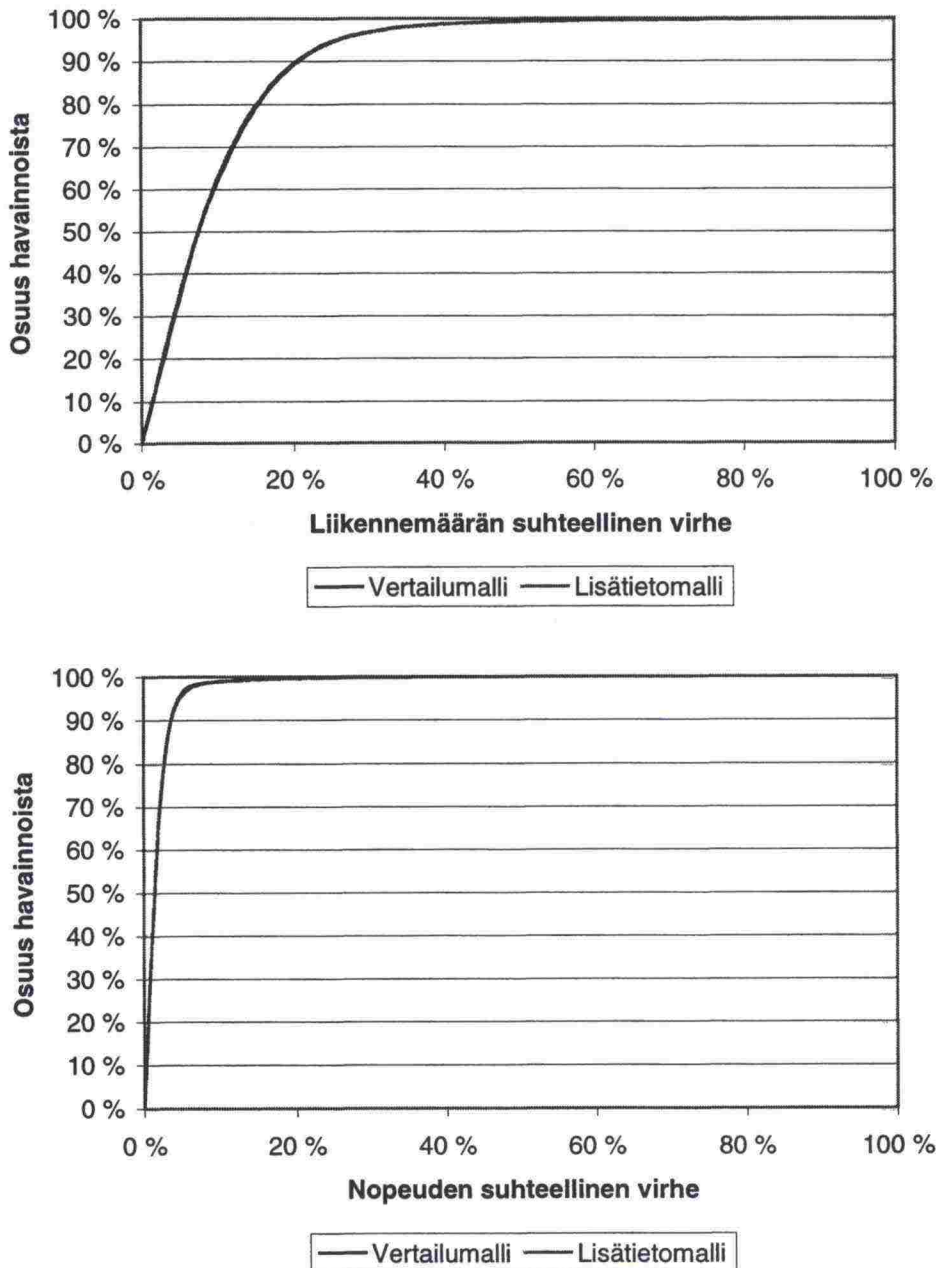
Vertailua varten sekä keskihajontatietoja hyödyntävistä malleista että niiden vertailumalleista valittiin skaalattuun aineistoon perustuvat liikennemäärän ennustamismallit sekä normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuvat nopeuden ennustamismallit. Malleille laskettiin kunkin virhetermin keskiarvo, joita verrattiin keskenään (taulukko 26). Nopeuskeskihajontatietoja hyödyntävät lisätietomallit osoittautuivat muutamien poikkeuksien paremmiksi kuin vertailumallit. Mallien väliset erot olivat suurelta osin suhteellisen pieniä. Osa eroista oli prosenteissa huomattavia, vaikkei suuria absoluuttisia eroja ollutkaan.

Taulukko 29. Normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuvien nopeuskeskihajontatietoja hyödyntävien mallien sekä niiden normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuvan liikennemäärän vertailuennustemallin ja skaalattuun aineistoon perustuvan keskinopeuden vertailuennustemallin keskimääräiset virhetermit ja niiden suhteelliset erot.

Virhetermi	Liikennemäärä			Nopeus		
	Lisätietomalli	Ero	Vertailumalli	Lisätietomalli	Ero	Vertailumalli
Keskimääräinen neliövirhe	43 000	1 %	44 000	6	0 %	6
Keskimääräinen absoluuttinen virhe	6	-1 271 %	0	0	83 %	0
Keskimääräinen absoluuttisen virheen itseisarvo	160	1 %	160	2	0 %	2
Keskimääräinen suhteellinen virhe	2 %	-21 %	2 %	0 %	251 %	0 %
Keskimääräinen suhteellisen virheen itseisarvo	10 %	0 %	10 %	2 %	0 %	2 %

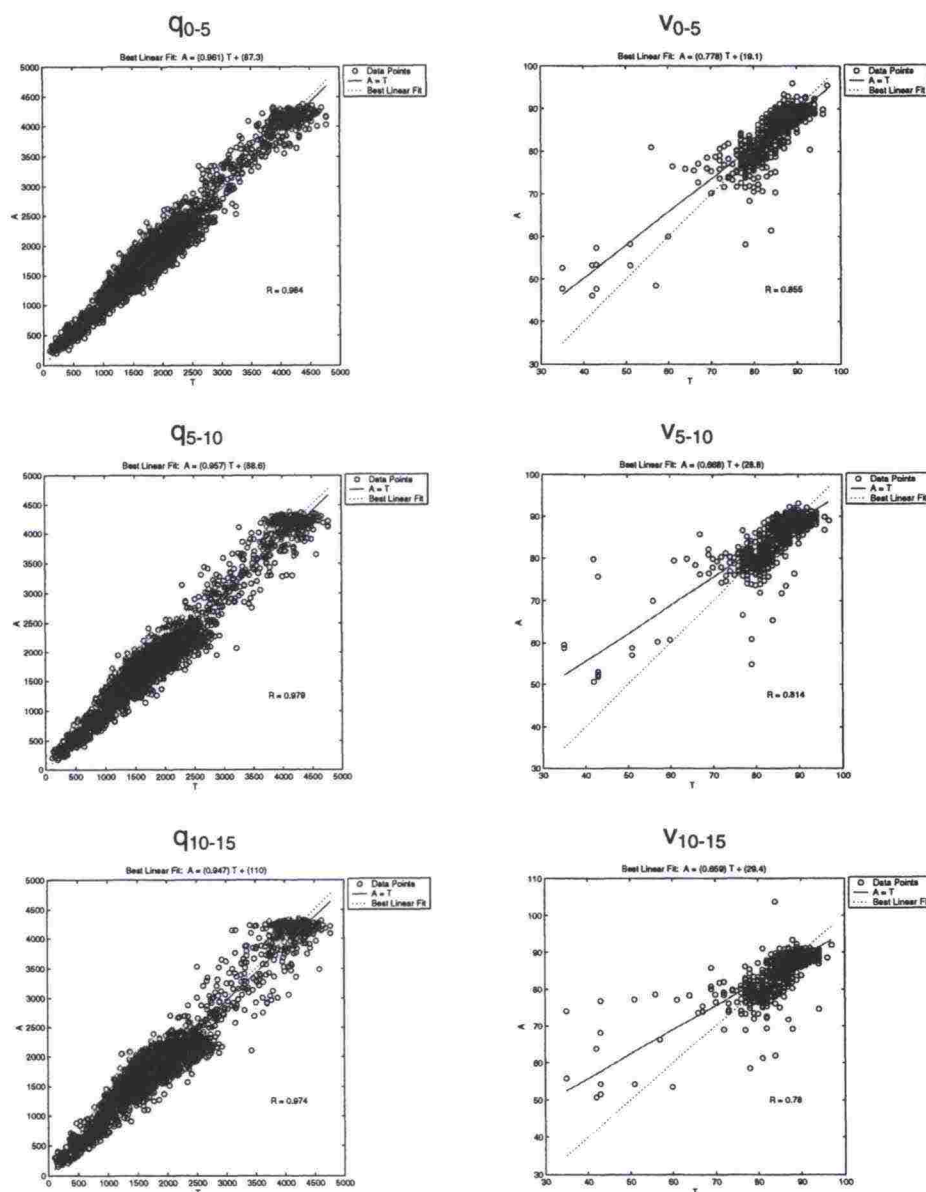
Kuvassa 16 on esitetty nopeuskeskihajontatietoja hyödyntävien mallien (lisätietomallit) ja niiden vertailumallien liikennemäärä- ja nopeusennusteiden suhteellisten virheiden kumulatiiviset summakäyrät. 90 prosentissa molempien mallien liikennemääräennusteista suhteellinen virhe oli korkeintaan 20 prosenttia. Molemmilla malleilla virhe oli korkeintaan kymmenen prosenttia 64 prosentissa ennusteista.

Molempien mallien nopeusennusteista 90 prosentissa suhteellinen virhe oli korkeintaan kolme prosenttia. Kummallakin mallilla 99 prosentissa nopeusennusteista suhteellinen virhe oli korkeintaan kymmenen prosenttia.



Kuva 17. Liikennemäärän ja keskinopeuden suhteellisten virheiden itseisarvojen summakäyrät nopeuskeskihajontatietoja hyödyntävälle mallille (lisätietomalli) ja sen vertailumallille. Liikennemääräennusteet perustuvat skaalattuun aineistoon ja nopeusennusteet normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon.

Kuvassa 18 on esitetty regressioanalyysi nopeuskeskihajontatietoja hyödyntävälle mallille. Kuvasta havaitaan, että vaikkei keskimääräisissä virhetermeissä ollutkaan eroja perusmalliin, nopeuskeskihajontatietoja hyödyntävä malli onnistui alhaisten nopeuksien ennustamisessa paremmin kuin perusmalli (esim. kuva 14).



Kuva 18.

Regressioanalyysi nopeuskeskihajontatietoja hyödyntävälle mallille. Liikennemäärän ennustamismalli käyttää skaalattua aineistoa ja keskinopeuden ennustamismalli normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa. X-akselilla on "oikea" ja y-akselilla ennustettu arvo. Liikennemäärää on merkitty q :lla ja keskinopeutta v :llä. Alaindeksit viittaavat ennustejaksoon.

7.3 Ennustejakson pituus

Ennustejakson pituuden vaikutusta ennustetuloksiin tutkittiin opettamalla neuroverkko ennustamaan liikennemäärää ja keskinopeus 30 minuuttia tarkasteluhestä eteenpäin. Malli opetettiin aineistolla, jossa liikenteen aamuhuipun havainnot olivat kahteen kertaan.

Taulukossa 30 on esitetty mallien virhetermit. Liikennemääräennusteissa normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuva malli antoi keskimääräisen neliövirheen ja keskimääräisen absoluuttisen virheen itseisarvolla mitattuna parempia ennusteita kuin skaalattuun aineistoon perustuva malli, joka oli puolestaan parempi keskimääräisen absoluuttisen ja suhteellisen virheen avulla mitattuna. Nopeusennusteissa skaalattuun aineistoon perustuva malli oli parempi kuin normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuva malli, kun paremmuutta mitattiin keskimääräisellä absoluuttisen virheen itseisarvolla tai keskimääräisen suhteellisen virheen itseisarvolla.

Voisi olettaa, että ennusteen laatu heikkenisi ennustejakson ajankohdan ollessa kauempana tarkasteluhetkestä. Näin ei kuitenkaan välttämättä ollut. Ne virhetermit, joissa tarkastellaan virheiden itseisarvoja (keskimääräinen absoluuttisen tai suhteellisen virheen itseisarvo sekä keskimääräinen neliövirhe), pienenevät ennustejakson ollessa yhä kauempana tarkasteluhetkestä (*taulukko 30*). Keskimääräinen absoluuttinen ja suhteellinen virhe kuitenkin käyttäytyivät oletuksen mukaisesti. Tästä voi tehdä johtopäätöksen, että kokonaisennustejakson pituuden kasvattaminen huononsi kaikkia ennusteita - eivätne ainoastaan uudet ennusteviisiminuuttiset olleet huonompia kuin lyhyemmän ennusteen antamat tulokset.

Taulukko 30. 30 minuuttia eteenpäin ennustavien mallien virhetermit. Taulukossa normeeratulla datalla tarkoitetaan normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa.

Virhetermi	Aikajakso	Liikennemäärä		Nopeus	
		Skaalattu	Normeerattu	Skaalattu	Normeerattu
Keskimääräinen neliövirhe	0 - 5	1 880 000	1 880 000	36	34
	5 - 10	1 890 000	1 880 000	33	34
	10 - 15	1 850 000	1 840 000	35	33
	15 - 20	1 850 000	1 830 000	32	33
	20 - 25	1 810 000	1 790 000	35	33
	25 - 30	1 790 000	1 790 000	31	34
Keskimääräinen absoluuttinen virhe	0 - 5	120	130	-1	-1
	5 - 10	120	135	0	0
	10 - 15	120	130	-1	-1
	15 - 20	130	140	0	-1
	20 - 25	140	150	-1	-1
	25 - 30	150	160	-1	-1
Keskimääräinen absoluuttisen virheen itseisarvo	0 - 5	1 030	1 020	3	3
	5 - 10	1 025	1 020	3	3
	10 - 15	1 010	1 010	3	3
	15 - 20	1 010	1 000	3	3
	20 - 25	1 000	990	3	3
	25 - 30	990	980	3	3
Keskimääräinen suhteellinen virhe	0 - 5	44 %	46 %	0 %	0 %
	5 - 10	43 %	45 %	0 %	0 %
	10 - 15	41 %	44 %	0 %	0 %
	15 - 20	42 %	42 %	0 %	0 %
	20 - 25	40 %	40 %	0 %	0 %
	25 - 30	38 %	39 %	0 %	0 %
Keskimääräinen suhteellisen virheen itseisarvo	0 - 5	81 %	82 %	4 %	4 %
	5 - 10	80 %	80 %	4 %	4 %
	10 - 15	78 %	78 %	4 %	4 %
	15 - 20	77 %	76 %	4 %	4 %
	20 - 25	74 %	74 %	4 %	4 %
	25 - 30	72 %	72 %	4 %	4 %

Regressioanalyysin tulokset vastasivat oletusta, että ennusteen laatu heikkenee ennusteajajakson ollessa yhä kauempana tarkasteluhetkestä (taulukko 31). Regressioanalyysin perusteella liikennemääräennusteista skaalattuun aineistoon perustuva malli oli parempi kuin normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuva malli kaikissa ennustejaksoissa. Nopeusennusteessa tilanne oli ensimmäisen 15 minuutin aikana päinvastainen, mutta jälkimmäisen 15 minuutin jakson aikana ennustemallien paremmuusjärjestys oli sama kuin liikennemääräennusteella.

Taulukko 31. 30 minuuttia eteenpäin ennustavien mallien regressioanalyysin tunnusluvut. Tunnuslukuista m tarkoittaa regressiosuoran kulma-kerrointa, b pistettä, jossa regressiosuora leikkaa y-akselin, ja r on selitysaste. Taulukossa normeeratulla datalla tarkoitetaan normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa.

Aikajakso	Tunnusluku	Liikennemäärä		Nopeus	
		Skaalattu	Normeerattu	Skaalattu	Normeerattu
0 - 5	m	0,97	0,96	0,69	0,75
	b	69	86	27	21
	r	0,98	0,98	0,88	0,88
5 - 10	m	0,96	0,96	0,65	0,71
	b	79	100	31	25
	r	0,98	0,98	0,84	0,85
10 - 15	m	0,96	0,95	0,68	0,69
	b	94	120	28	27
	r	0,98	0,97	0,83	0,82
15 - 20	m	0,95	0,94	0,68	0,61
	b	100	140	28	34
	r	0,97	0,97	0,82	0,79
20 - 25	m	0,97	0,92	0,68	0,57
	b	120	180	28	37
	r	0,97	0,96	0,80	0,76
25 - 30	m	0,93	0,91	0,62	0,53
	b	140	190	33	40
	r	0,96	0,95	0,76	0,73

30 minuutin ennustemallille tehtiin vertailumalli, joka oli muuten aivan samanlainen, mutta se teki ennusteen ainoastaan 15 minuuttia tarkasteluhetkestä eteenpäin. Vertailumalli oli siis jaettu kahteen osamalliin, joista toinen ennusti liikennemäärää ja toinen keskinopeutta, mallissa oli yhtä monta piiloneuronia kuin 30 minuutin mallissa ja se perustui samaan opetusdataan. Näin erot neuroverkon rakenteessa tai opetusdatassa eivät päässeet vaikuttamaan vertailun tuloksiin.

On huomattava, ettei tuloksia kannata verrata luvussa 6 esitettyjen mallien tuloksiin, koska niissä voitiin käyttää syötetietojen vähemmyydestä johtuen useampia piiloneuroneja kuin nyt tehdyissä malleissa. Nyt tehtyjen mallien tulokset saataisiin lähemmäksi kappaleen 6 mallien tuloksia, jos opetusdataa olisi enemmän, jolloin piiloneuronien määrää voitaisiin lisätä. Ainoastaan samaan opetusdataan ja samanlaiseen piilokerrokseen perustuvat mallit ovat siis vertailukelpoisia keskenään.

Taulukossa 32 on esitetty vertailumallin virheet ja taulukossa 33 regressioanalyysin tulokset. Liikennemääräennusteissa skaalattuun aineistoon perustuva malli oli parempi kuin normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuva malli, kun paremmuutta mitattiin keskimääräisellä absoluuttisella tai suhteellisella virheellä. Kun mittarina oli keskimääräinen neliövirhe tai absoluuttisen virheen itseisarvo, tilanne oli toisin päin. Nopeusennusteissa

tilanne oli muuten aivan vastaava, mutta keskimääräisen absoluuttisen virheen itseisarvolla mitattuna mallien välillä ei ollut eroa. Skaalattuun aineistoon perustuvat mallit menestyivät regressioanalyysissä paremmin kuin normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuvat ennustemallit.

Taulukko 32. 30 minuuttia eteenpäin ennustavien mallien 15 minuutin vertailumallin virhetermit. Taulukossa normeeratulla aineistolla tarkoitetaan normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa.

Virhetermi	Aikajakso	Liikennemäärä		Nopeus	
		Skaalattu	Normeerattu	Skaalattu	Normeerattu
Keskimääräinen neliövirhe	0 - 5	1 830 000	1 820 000	33	30
	5 - 10	1 820 000	1 810 000	33	31
	10 - 15	1 810 000	1 800 000	34	31
Keskimääräinen absoluuttinen virhe	0 - 5	84	85	0	-1
	5 - 10	87	89	0	-1
	10 - 15	89	89	0	-1
Keskimääräinen absoluuttisen virheen itseisarvo	0 - 5	1 020	1 010	3	3
	5 - 10	1 010	1 000	3	3
	10 - 15	1 010	1 000	3	3
Keskimääräinen suhteellinen virhe	0 - 5	41 %	43 %	0 %	-1 %
	5 - 10	40 %	42 %	0 %	-1 %
	10 - 15	38 %	40 %	0 %	-1 %
Keskimääräinen suhteellisen virheen itseisarvo	0 - 5	79 %	79 %	4 %	4 %
	5 - 10	78 %	78 %	4 %	4 %
	10 - 15	76 %	76 %	4 %	4 %

Taulukko 33. 30 minuuttia eteenpäin ennustavien mallien 15 minuutin vertailumallin regressioanalyysin tunnusluvut. Tunnusluvuista m tarkoittaa regressiosuoran kulmakerrointa, b pistettä, jossa regressiosuora leikkaa y -akselin, ja r on selitysaste. Taulukossa normeeratulla datalla tarkoitetaan normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa.

Aikajakso	Tunnusluku	Liikennemäärä		Nopeus	
		Skaalattu	Normeerattu	Skaalattu	Normeerattu
0 - 5	m	0,97	0,97	0,75	0,69
	b	63	71	22	26
	r	0,98	0,98	0,88	0,87
5 - 10	m	0,96	0,96	0,67	0,57
	b	73	86	29	37
	r	0,98	0,98	0,84	0,81
10 - 15	m	0,96	0,95	0,68	0,53
	b	88	100	28	41
	r	0,98	0,97	0,82	0,80

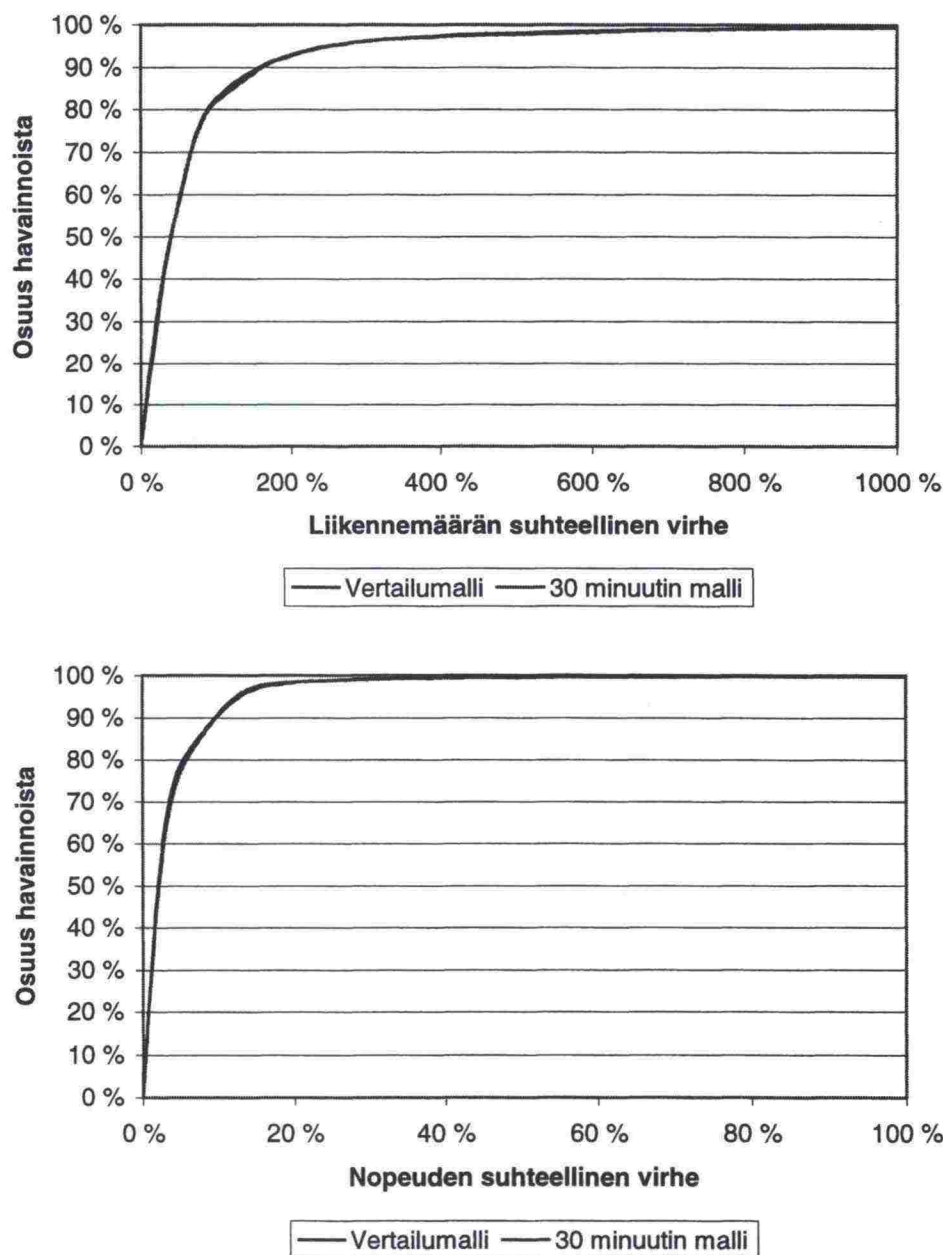
Vertailuun valittiin molemmista malleista skaalattuun aineistoon perustuvat osamallit. Näille laskettiin kunkin virhetermin keskiarvo, joita verrattiin keskenään (taulukko 34). 30 minuutin ennustemallin virhetermit olivat suurempia kuin ver-

tailumallin. On kuitenkin huomattava, että keskimääräisen suhteellisen virheen itseisarvolla mitattuna 30 minuutin malli antoi keskimäärin hieman parempia tuloksia kuin vertailumalli. Mallien väliset erot olivat suurelta osin suhteellisen pieniä. Osa eroista oli prosenteissa huomattavia, vaikkei eroissa suuria absoluuttisia eroja ollutkaan.

Taulukko 34. Skaalattuun aineistoon perustuvan 30 minuuttia eteenpäin ennustavan mallin ja sen 15 minuutin vertailumallin keskimääräiset virhetermit ja niiden suhteelliset erot.

Virhetermi	Liikennemäärä			Nopeus		
	30 minuutin malli	Ero	Vertailumalli	30 minuutin malli	Ero	Vertailumalli
Keskimääräinen neliövirhe	1 850 000	-2 %	1 820 000	34	-1 %	33
Keskim. absoluuttinen virhe	130	-52 %	87	-1	-39 %	0
Keskimääräinen absoluuttisen virheen itseisarvo	1 010	0 %	1 010	3	-2 %	3
Keskimääräinen suhteellinen virhe	41 %	-4 %	40 %	0 %	-84 %	0 %
Keskimääräinen suhteellisen virheen itseisarvo	77 %	1 %	78 %	4 %	-2 %	4 %

Kuvassa 19 on esitetty skaalattuun aineistoon perustuvan 30 minuuttia eteenpäin ennustavan mallin ja sen skaalattuun aineistoon perustuvan vertailumallin suhteellisten virheiden summakäyrät. 90 prosentissa 30 minuuttia eteenpäin ennustavan mallin liikennemääräennusteista suhteellinen virhe oli korkeintaan 166 prosenttia ja vertailumallin liikennemääräennusteista 164 prosenttia. 30 minuuttia eteenpäin ennustavan mallin liikennemääräennusteista 17 prosentissa ja vertailumallin liikennemääräennusteista 15 prosentissa suhteellinen virhe oli korkeintaan kymmenen prosenttia. Molempien mallien nopeusennusteista 92 prosentissa suhteellinen virhe oli korkeintaan kymmenen prosenttia.



Kuva 19. Liikennemäärän ja keskinopeuden suhteellisten virheiden itseisarvojen summakäyrät skaalattuun aineistoon perustuvalle 30 minuuttia eteenpäin ennustavalle mallille ja sen skaalattuun aineistoon perustuvalle vertailumallille.

7.4 Poikkileikkausten sijainti

7.4.1 Kaksi poikkileikkausta

Ennustemalleja tehtäessä aineistoa ei aina ole käytettävissä niin monelta poikkileikkaukselta kuin edellä. Poikkileikkausten keskinäisen sijainnin vaikutusta ennusteen laatuun tutkittiin olettamalla, että käytössä on aineistoa ainoastaan

kahdelta poikkileikkaukselta eli ennuste- ja yhdeltä lisätietopoikkileikkaukselta. Ennusteet tehtiin edelleen poikkileikkaukselle 11. Käytettävissä olevista lisätietopoikkileikkauksista poikkileikkaus 10 sijaitsi 420 metriä, poikkileikkaus yhdeksän 807 metriä, poikkileikkaus kahdeksan 1 160 metriä ja poikkileikkaus kuusi 1 878 metriä ennen ennustepoikkileikkausta. Poikkileikkaus 12 sijaitsi 392 metriä ja poikkileikkaus neljatoista 1 080 metriä ennustepoikkileikkauksen jälkeen (kuva 1).

Ensin tehtiin ennustemallit, jotka perustuivat poikkileikkausten 11 ja 6 skaalatun tai normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon. Mallit koostuivat yksittäisestä neuroverkosta, joka teki ennusteen sekä liikennemäärälle että keskinopeudelle. Ennustejaksona oli 15 minuuttia ja ennuste perustui edellisen 15 minuutin liikennemäärä- ja keskinopeustietoihin kyseisiltä poikkileikkauksilta.

Skaalattuun aineistoon perustuvassa mallissa oli siis 12 syöteneuronia, jolloin piiloneuronien määräksi saatiin 8. Normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuvassa mallissa syöteneuroneja oli 8 ja piiloneuroneja 10. Mallien tuloksista lasketut virhetermit on esitetty taulukossa 35. Taulukosta havaitaan, että nopeusennusteiden keskimääräistä neliövirhettä lukuun ottamatta skaalattuun aineistoon perustuva malli teki parempia ennusteita kuin normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuva malli. Loput kahden poikkileikkauksen liikennetietoihin perustuvat ennustemallit tehtiin skaalattuun aineistoon perustuviksi.

Taulukko 35. Poikkileikkausten 11 ja 6 aineistoon perustuvien ennustemallien virhetermit. Taulukossa normeeratulla aineistolla tarkoitetaan normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa.

Virhetermi	Aikajakso	Liikennemäärä		Nopeus	
		Skaalattu	Normeerattu	Skaalattu	Normeerattu
Keskimääräinen neliövirhe	0 - 5	37 000	44 000	5	5
	5 - 10	45 000	54 000	6	6
	10 - 15	53 000	66 000	7	6
Keskimääräinen absoluuttinen virhe	0 - 5	-7	26	0	0
	5 - 10	-6	19	0	0
	10 - 15	0	32	0	0
Keskimääräinen absoluuttisen virheen itseisarvo	0 - 5	150	170	1	2
	5 - 10	160	180	2	2
	10 - 15	180	200	2	2
Keskimääräinen suhteellinen virhe	0 - 5	0 %	4 %	0 %	0 %
	5 - 10	0 %	4 %	0 %	0 %
	10 - 15	1 %	5 %	0 %	0 %
Keskimääräinen suhteellisen virheen itseisarvo	0 - 5	9 %	11 %	2 %	2 %
	5 - 10	10 %	11 %	2 %	2 %
	10 - 15	11 %	13 %	2 %	2 %

Kaikkien kahden poikkileikkauksen liikennetietoja lähtötietoinaan käyttävien ennustemallien virhetermit on esitetty *taulukossa 36*. Taulukon perusteella malleja ei ole helppo asettaa paremmuusjärjestykseen, koska yksikään malleista ei ollut selvästi parempi kuin muut, vaan paremmuusjärjestys vaihteli muuttujan ja virhetermin mukaan. Kahden poikkileikkauksen liikennetietoihin perustuvien mallien paremmuusjärjestystä selvitettiin pisteyttämällä kunkin muuttujan kukin virhetermi mallien kesken siten, että parhaasta tuloksesta (= pienin virhetermi) sai 5 pistettä, toiseksi parhaasta 4 ja niin edelleen siten, että huonoin tulos jäi pisteittä. Tulokseksi saadut pisteet on annettu *taulukossa 37*. Ennen ennustepoikkileikkausta sijaitsevia lisätietopoikkileikkauksia käyttävistä malleista paras oli poikkileikkauksiin 11 ja 6 perustuva malli. Mallien suoritus huononi sitä mukaa, kun lisätietopoikkileikkaus lähestyi ennustepoikkileikkausta. Ennustepoikkileikkauksen jälkeen sijaitsevia poikkileikkauksia käyttävillä malleilla tilanne oli sitä vastoin päinvastainen.

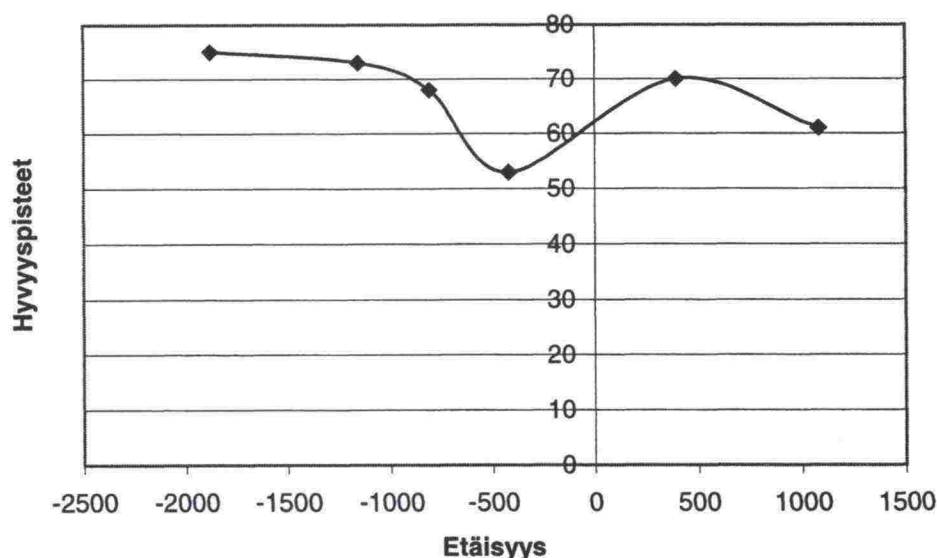
Taulukko 36. Kahden poikkileikkauksen skaalattuun aineistoon perustuvien ennustemallien virhetermit.

Virhetermi	Muuttuja	Poikkileikkaukset					
		11 ja 6	11 ja 8	11 ja 9	11 ja 10	11 ja 12	11 ja 14
Keskimääräinen neliövirhe	Q0-5	37 000	37 000	39 000	42 000	45 000	42 000
	Q5-10	45 000	44 000	48 000	51 000	55 000	51 000
	Q10-15	54 000	52 000	58 000	62 000	68 000	60 000
	V0-5	5	5	5	5	5	5
	V5-10	6	7	6	7	6	7
	V10-15	7	7	7	7	7	9
Keskimääräinen absoluuttinen virhe	Q0-5	-7	3	6	-2	20	12
	Q5-10	-6	-1	-2	0	20	-24
	Q10-15	0	3	-4	4	24	-29
	V0-5	0	0	0	0	0	0
	V5-10	0	0	0	0	0	0
	V10-15	0	0	0	0	0	0
Keskimääräinen absoluuttisen virheen itseisarvo	Q0-5	150	150	150	160	160	160
	Q5-10	160	160	170	175	180	170
	Q10-15	180	175	180	190	190	190
	V0-5	1	1	1	2	2	2
	V5-10	2	2	2	2	2	2
	V10-15	2	2	2	2	2	2
Keskimääräinen suhteellinen virhe	Q0-5	0 %	2 %	1 %	3 %	4 %	4 %
	Q5-10	0 %	2 %	1 %	4 %	4 %	1 %
	Q10-15	1 %	2 %	1 %	4 %	4 %	1 %
	V0-5	0 %	0 %	0 %	0 %	0 %	0 %
	V5-10	0 %	0 %	0 %	0 %	0 %	0 %
	V10-15	0 %	0 %	0 %	0 %	0 %	0 %
Keskimääräinen suhteellisen virheen itseisarvo	Q0-5	9 %	9 %	9 %	10 %	10 %	10 %
	Q5-10	10 %	10 %	10 %	12 %	11 %	10 %
	Q10-15	11 %	11 %	11 %	12 %	12 %	11 %
	V0-5	2 %	2 %	2 %	2 %	2 %	2 %
	V5-10	2 %	2 %	2 %	2 %	2 %	2 %
	V10-15	2 %	2 %	2 %	2 %	2 %	2 %

Taulukko 37. Taulukon 36 tuloksista kahden poikkileikkauksen aineistoihin perustuvilla malleille annetut hyvyyspisteet ja lisätietopoikkileikkauksen etäisyys ennustepoikkileikkauksesta (negatiivinen etäisyys tarkoittaa sitä, että poikkileikkaus sijaitsee ennen ennustepoikkileikkausta).

Poikkileikkaukset, joiden tietoihin malli perustuu	Hyvyyspisteet	Etäisyys ennustepoikkileikkauksesta
11 ja 6	75	-1878
11 ja 8	73	-1160
11 ja 9	68	-807
11 ja 10	53	-420
11 ja 12	70	392
11 ja 14	61	1080

Tuloksia on havainnollistettu kuvassa 20, jossa hyvyyspisteet on piirretty etäisyyden funktiona. Kuvan perusteella vaikuttaa siltä, että jos käytettävissä on liikennetiedot ainoastaan kahdelta poikkileikkaukselta, lisätietopoikkileikkaus kannatti sijoittaa vähintään 800 metriä ennen ennustepoikkileikkausta. Jos lisätietopoikkileikkaus on sijoitettava ennustepoikkileikkauksen jälkeen, niiden välisen etäisyyden kannatti olla pienempi, noin 400 metriä. Tässä on huomattava, että optimisijainnit ovat voimakkaasti riippuvaisia kulloisestakin kohteesta.



Kuva 20. Mallin hyvyyspisteet lisätietopoikkileikkauksen etäisyyden funktiona. Ennustepoikkileikkaus sijaitsee origossa.

7.4.2 Kolme poikkileikkausta

Toisena asiana selvitettiin sitä, mihin kolmas poikkileikkaus kannattaa sijoittaa. Lähtökohtana pidettiin mallia, joka sai tiedot poikkileikkauksilta 11 ja 6, joiden välinen etäisyys oli noin 1 900 metriä. Eri malleille lasketut virhetermit on esitetty taulukossa 38. Tulokset on pisteytetty vastaavasti kuin edellä siten, että parhaasta tuloksesta (= pienin virhetermi) sai 4 pistettä, toiseksi parhaasta 3 pis-

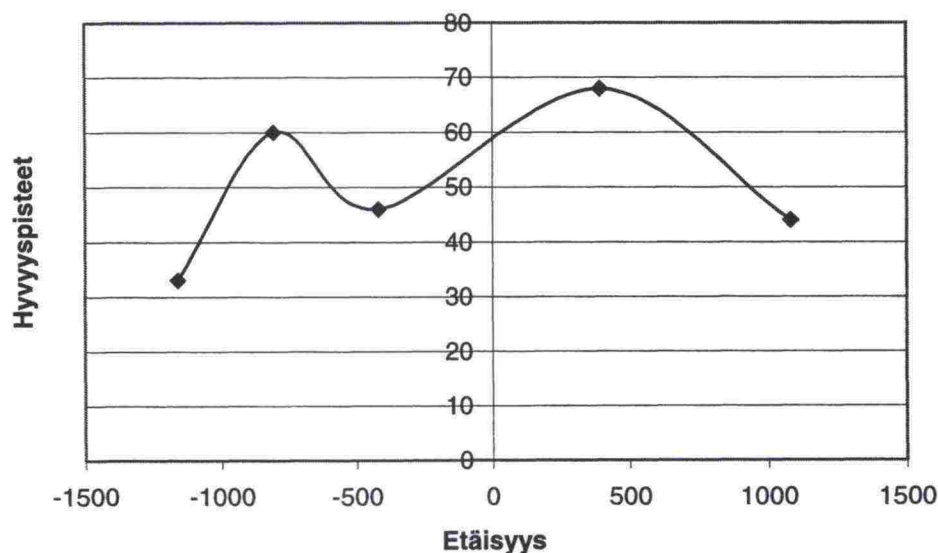
tettä ja niin edelleen. Pisteet on annettu *taulukossa 39* ja niitä on havainnollistettu *kuvassa 21*. Kuvasta ja taulukosta havaitaan, että kun ennustemalli sai liikennetiedot ennustepoikkileikkauksen lisäksi poikkileikkauksesta, joka sijaitsi noin 1 900 metriä ennen ennustepoikkileikkausta, toiseksi lisätietopoikkileikkaukseksi kannatti valita poikkileikkaus, joka sijaitsi noin 400 metriä ennustepoikkileikkauksen jälkeen. Toiseksi paras valinta oli poikkileikkaus, joka sijaitsi noin 800 metriä ennen ennustepoikkileikkausta. Tässäkin on huomattava, että näiden optimaalisten lisätietopoikkileikkausten sijainti on voimakkaasti riippuvainen kul-loisestakin kohteesta.

Taulukko 38. Kolmen poikkileikkauksen liikennetietoihin perustuvien ennustemallien virhetermit.

Virhetermi	Muuttuja	Poikkileikkaukset				
		11, 6 ja 8	11, 6 ja 9	11, 6 ja 10	11, 6 ja 12	11, 6 ja 14
Keskimääräinen nettovirhe	Q ₀₋₅	36 000	36 000	38 000	37 000	37 000
	Q ₅₋₁₀	45 000	46 000	44 000	46 000	45 000
	Q ₁₀₋₁₅	54 000	56 000	54 000	56 000	53 000
	V ₀₋₅	6	6	5	7	5
	V ₅₋₁₀	6	10	6	7	7
	V ₁₀₋₁₅	7	16	7	7	8
Keskimääräinen absoluuttinen virhe	Q ₀₋₅	1	8	-1	2	-4
	Q ₅₋₁₀	-4	5	-4	5	-12
	Q ₁₀₋₁₅	-1	7	-1	12	-12
	V ₀₋₅	0	0	0	0	0
	V ₅₋₁₀	0	0	0	0	0
	V ₁₀₋₁₅	0	0	0	0	0
Keskimääräinen absoluuttisen virheen itseisarvo	Q ₀₋₅	150	150	150	150	150
	Q ₅₋₁₀	160	170	160	170	160
	Q ₁₀₋₁₅	180	180	180	180	180
	V ₀₋₅	2	1	2	2	2
	V ₅₋₁₀	2	2	2	2	2
	V ₁₀₋₁₅	2	2	2	2	2
Keskimääräinen suhteellinen virhe	Q ₀₋₅	2 %	1 %	1 %	2 %	1 %
	Q ₅₋₁₀	2 %	1 %	2 %	2 %	1 %
	Q ₁₀₋₁₅	2 %	2 %	3 %	3 %	1 %
	V ₀₋₅	0 %	0 %	0 %	0 %	0 %
	V ₅₋₁₀	0 %	0 %	0 %	0 %	0 %
	V ₁₀₋₁₅	0 %	0 %	0 %	0 %	0 %
Keskimääräinen suhteellisen virheen itseisarvo	Q ₀₋₅	9 %	9 %	9 %	9 %	9 %
	Q ₅₋₁₀	10 %	10 %	10 %	10 %	10 %
	Q ₁₀₋₁₅	11 %	11 %	11 %	11 %	11 %
	V ₀₋₅	2 %	2 %	2 %	2 %	2 %
	V ₅₋₁₀	2 %	2 %	2 %	2 %	2 %
	V ₁₀₋₁₅	2 %	2 %	2 %	2 %	2 %

Taulukko 39. Taulukon 38 tuloksista annetut hyvyyspisteet ja toisen lisätietopoikkileikkauksen etäisyys ennustepoikkileikkauksesta (negatiivinen etäisyys tarkoittaa sitä, että poikkileikkaus sijaitsee ennen ennustepoikkileikkausta). Pisteiden 6 ja 11 välinen etäisyys on noin 1 900 metriä.

Poikkileikkaukset, joiden tietoihin malli perustuu	Hyvyyspisteet	Kolmannen poikkileikkauksen etäisyys ennustepoikkileikkauksesta (m)
11, 6 ja 8	33	-1160
11, 6 ja 9	60	-807
11, 6 ja 10	46	-420
11, 6 ja 12	68	392
11, 6 ja 14	44	1080



Kuva 21. Hyvyyspisteet toisen lisätietopoikkileikkauksen etäisyyden funktiona. Ennustepoikkileikkaus sijaitsee origossa ja ensimmäinen lisätietopoikkileikkaus noin 1 900 metriä ennustepoikkileikkauksesta ylävirtaan (-1 900 m).

7.4.3 Lisätietopoikkileikkauksilla saavutettava etu

Lisätietopoikkileikkauksia harkittaessa kysymys tietenkin kuuluu, saadaanko lisämittauspisteen liikennetietojen avulla tehtyä tarkempia ennusteita. Asiaa selvitettiin tarkastellussa kohteessa tekemällä ainoastaan ennustepoikkileikkauksen liikennetietoihin perustuva ennustemalli. Mallin ennusteiden hyvyttä kuvaavat virhetermit on esitetty taulukossa 40.

Taulukko 40. Pelkän ennustepoikkileikkauksen (11) liikennetietoihin perustuvan ennustemallin virhetermit.

Muuttuja	Virhetermi				
	Keskimääräinen neliövirhe	Keskimääräinen absoluuttinen virhe	Keskimääräinen absoluuttisen virheen itseisarvo	Keskimääräinen suhteellinen virhe	Keskimääräinen suhteellisen virheen itseisarvo
q ₀₋₅	44 000	3	160	2 %	10 %
q ₅₋₁₀	54 000	0	170	2 %	10 %
q ₁₀₋₁₅	69 000	4	190	2 %	11 %
v ₀₋₅	5	0	2	0 %	2 %
v ₅₋₁₀	6	0	2	0 %	2 %
v ₁₀₋₁₅	7	0	2	0 %	2 %

Pelkän ennustepoikkileikkauksen liikennetietoihin perustuvan mallin tuloksia verrattiin yhden ja kahden lisätietopoikkileikkauksen liikennetiedot syötteinä saaneiden mallien tuloksiin. Lisätietomalleiksi valittiin kunkin malliryhmän malleista parhaat (syötteinä poikkileikkausten 11 ja 6 sekä 11, 6 ja 12 liikennetiedot). Tulokset pisteytettiin siten, että parhaasta tuloksesta sai kaksi pistettä, keskimmäisestä yhden ja huonoin jäi pisteittä. Tällä tavalla pisteytettynä malli, joka perustui ainoastaan ennustepoikkileikkauksen liikennetietoihin, sai 23 pistettä, yhden lisätietopoikkileikkauksen malli 46 pistettä ja kahden lisätietopoikkileikkauksen malli 21 pistettä.

Tämän yksinkertaisen tarkastelun perusteella vaikuttaa siis siltä, että yhdellä lisätietopoikkileikkauksella päästiin keskimääräisesti tarkasteltuna parhaisiin tuloksiin ja ettei toisen lisätietopoikkileikkauksen mukaan ottaminen kannattanut. On kuitenkin huomattava, että mitä kauempana mittauspoikkileikkaukset sijaitsevat häiriön tapahtumapaikasta tai ruuhkan syntypisteestä, sitä hitaammin malli voi reagoida liikennetilannemuutoksiin. Tällä perusteella useamman lisätietopoikkileikkauksen mukaan ottaminen on perusteltua.

8 Naiivi vertailumalli

Naiiviksi vertailumalliksi tehtiin malli, joka perustui liikennekäyttäytymisen säännönmukaisuuteen. Mallissa kolmelle eri päivätyypille (arkipäivä, lauantai ja sunnuntai/pyhä) määriteltiin keskimääräinen liikennemäärä ja keskinopeus kullekin viiden minuutin jaksolle. Mallille käytettiin vertailumallina kahteen osamalliin jaettua perusmallia, jossa liikennemääräennuste perustui skaalattuun ja nopeusennuste normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon.

Taulukossa 41 on esitetty mallien virhetermit. Taulukosta nähdään, että naiivin mallin tekemien liikennemääräennusteiden absoluuttiset virheet olivat pienempiä kuin perusmallin. Suhteellisia virheitä tarkasteltaessa paremmuusjärjestys oli päinvastainen. Ensimmäisen ennustejakson osalta perusmalli teki pienemmän keskimääräisen neliövirheen, mutta jälkimmäisillä jaksoilla naiivi malli oli parempi. Nopeusennusteissa perusmalli oli naiivia mallia parempi keskimääräisellä neliövirheellä ja suhteellisilla virheillä mitattuna. Absoluuttisilla virheillä mitattuna naiivi malli oli parempi, mutta ero oli hyvin pieni.

Taulukko 41. Naiivin ennustemallin ja kahteen osamalliin jaetun perusmallin virhetermit (liikennemääräennuste perustuu skaalattuun ja nopeusennuste normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon).

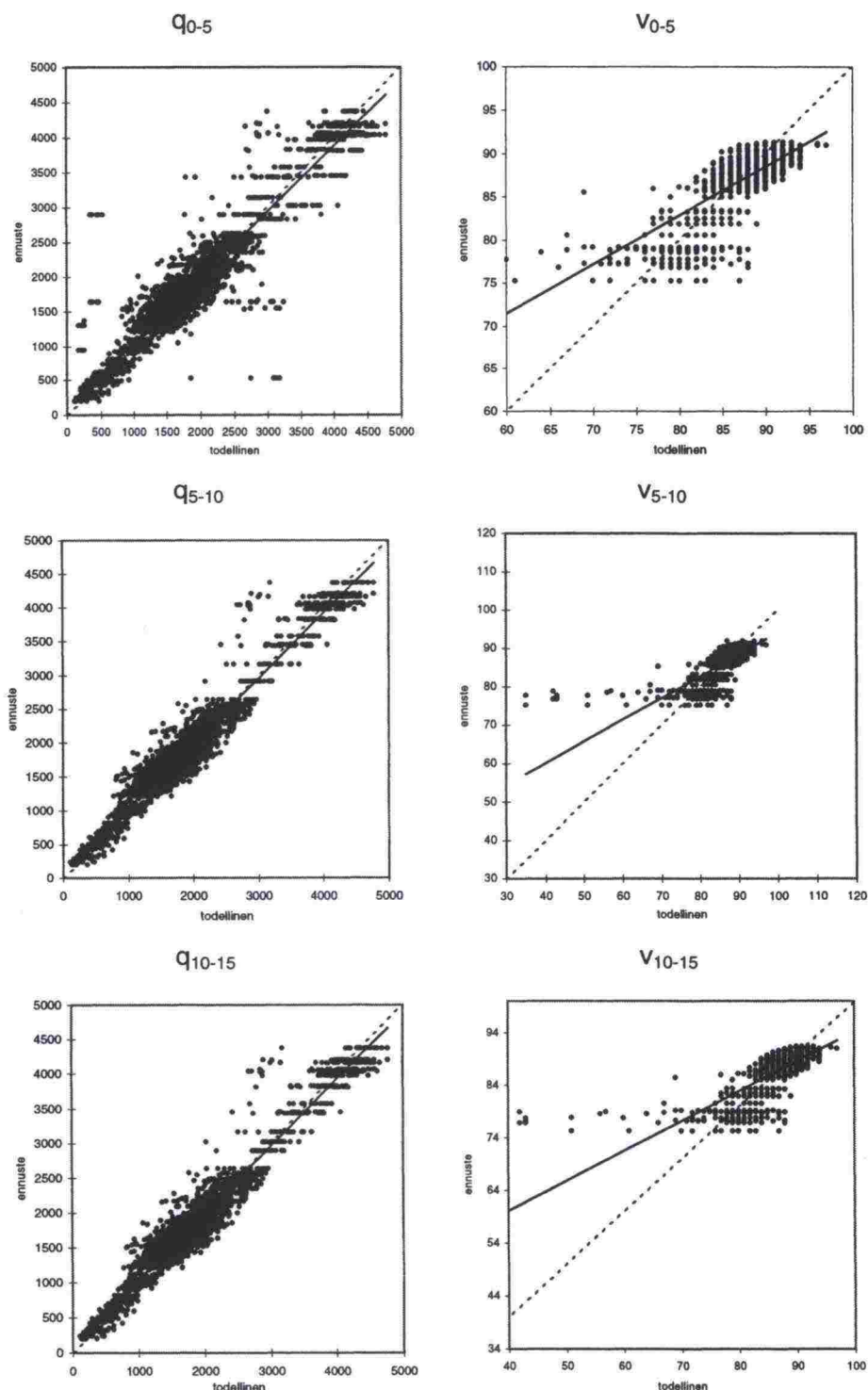
Virhetermi	Aikajakso	Liikennemäärä		Nopeus	
		Perusmalli	Naiivi malli	Perusmalli	Naiivi malli
Keskimääräinen neliövirhe	0 – 5	36 000	40 000	5	8
	5 – 10	44 000	40 000	6	8
	10 – 15	51 000	40 000	7	8
Keskimääräinen absoluuttinen virhe	0 – 5	-11	0	0	0
	5 – 10	-18	0	0	0
	10 – 15	-16	0	0	0
Keskimääräinen absoluuttisen virheen itseisarvo	0 – 5	150	150	1	2
	5 – 10	160	150	1	2
	10 – 15	170	150	2	2
Keskimääräinen suhteellinen virhe	0 – 5	-1 %	1 %	0 %	0 %
	5 – 10	-1 %	1 %	0 %	0 %
	10 – 15	-1 %	2 %	0 %	0 %
Keskimääräinen suhteellisen virheen itseisarvo	0 – 5	8 %	9 %	2 %	2 %
	5 – 10	8 %	9 %	2 %	2 %
	10 – 15	9 %	9 %	2 %	2 %

Kuvassa 22 on esitetty naiiville mallille tehty regressioanalyysi. Analyysin tunnuslukuja on verrattu perusmallin vastaaviin tunnuslukuihin taulukossa 42. Kuvasta havaitaan, että naiivin mallin liikennemääräennuste oli parempi kuin nopeusennuste. Liikennemääräennusteellakin oli tosin taipumus olla välillä reilusti liian suuri. Nopeusennusteista havaitaan, ettei naiivi malli kyennyt ennustamaan ruuhkatilanteiden erittäin alhaisia nopeuksia eli se ei kyennyt vastaamaan dynaamisiin satunnaisiin muutoksiin. Naiivin mallin tekemät alimmat nopeusen-

nusteet olivat suuruusluokkaa 75 km/h. Taulukosta havaitaan, että perusmalli teki regressioanalyysin tunnusluvuilla mitattuna parempia liikennemääräennusteita ensimmäiselle ennustejaksolle, mutta naiivin mallin tekemät liikennemääräennusteet olivat seuraaville ennustejaksoille hieman parempia. Perusmalli antaa sitä vastoin regressioanalyysin tunnusluvuilla mitattuna huomattavasti parempia nopeusennusteita kuin naiivi malli. Saaduista tuloksista voi tehdä johtopäätöksen, että liikennemäärä oli suurena säännönmukaisempi kuin nopeus, eli naiivi malli kykeni tekemään kilpailukykyisiä liikennemääräennusteita, mutta dynaamisten nopeusennusteiden tekemiseen kannatti käyttää neuroverkkoa.

Taulukko 42. Naiivin ennustemallin perusmallien regressioanalyysin tunnusluvut. Tunnusluvuista m tarkoittaa regressiosuoran kulmakerrointa, b pistettä, jossa regressiosuora leikkaa y-akselin, ja r on selitystaste.

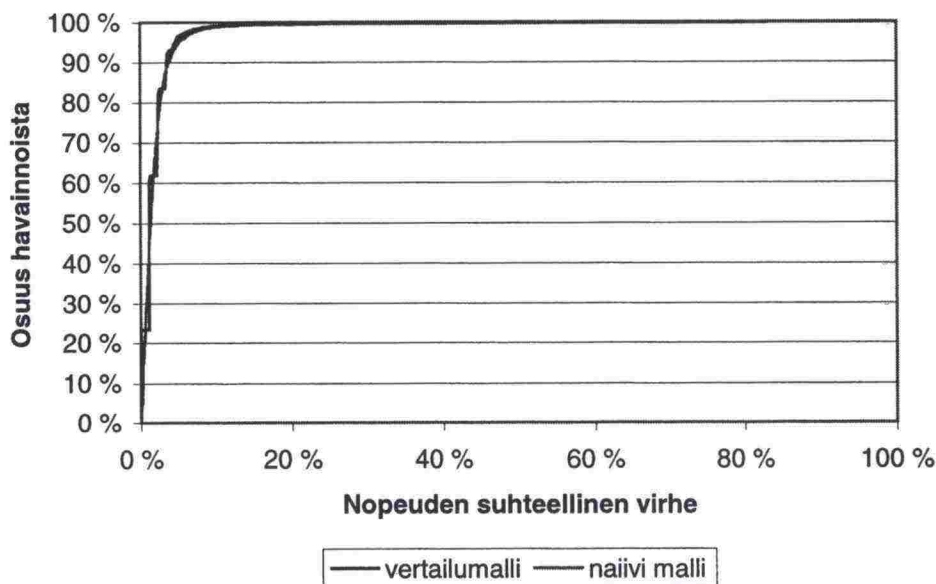
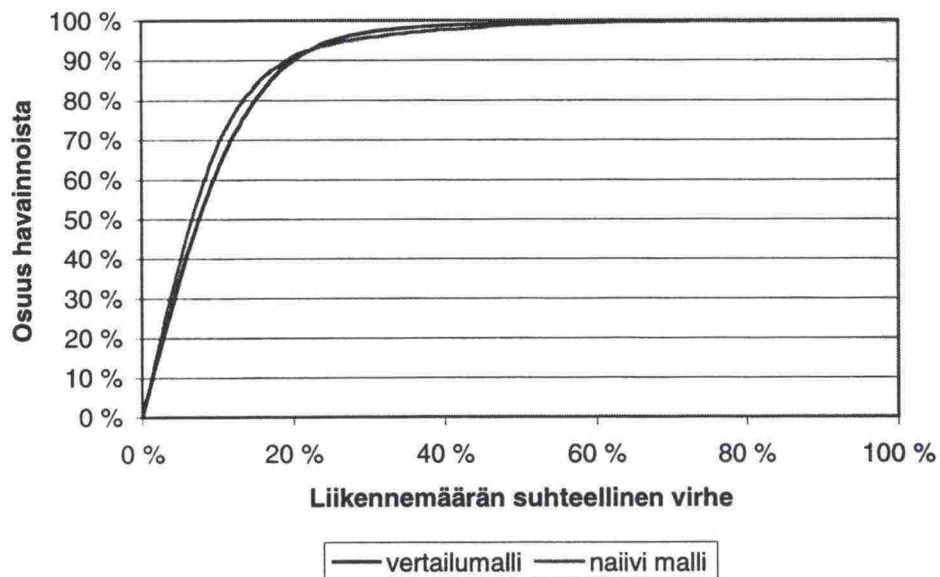
Aikajakso	Tunnusluku	Liikennemäärä		Nopeus	
		Perusmalli	Naiivi malli	Perusmalli	Naiivi malli
0 – 5	m	0,95	0,94	0,85	0,57
	b	91	122	13	37
	r	0,98	0,96	0,85	0,75
5 – 10	m	0,94	0,96	0,73	0,57
	b	90	83	23	37
	r	0,97	0,98	0,81	0,75
10 - 15	m	0,95	0,96	0,73	0,57
	b	114	84	24	37
	r	0,97	0,98	0,78	0,75



Kuva 22. Regressioanalyysi naiville mallille. Kuvissa x-akselilla on mitatut arvot ja y-akselilla naiivin mallin ennustamat arvot.

Kuvassa 23 on esitetty naiivin mallin ja sen vertailumallin suhteellisten virheiden kumulatiiviset summakäyrät. Molemmilla malleilla 90 prosentissa liikennemääräennusteista suhteellinen virhe oli korkeintaan 20 prosenttia. 70 prosentissa naiivin mallin liikennemääräennusteista suhteellinen virhe oli korkeintaan kymmenen prosenttia. Vertailumallilla vastaava luku oli 64 prosenttia. Molempien mallien

nopeusennusteista 90 prosentissa suhteellinen virhe oli korkeintaan neljä prosenttia ja 99 prosentissa korkeintaan kymmenen prosenttia.



Kuva 23.

Liikennemäärän ja keskinopeuden suhteellisten virheiden itseisarvojen kumulatiiviset summakäyrät naiiville mallille ja kahteen osamalliin jaetulle mallille (vertailumalli), jossa liikennemääräennusteet perustuvat skaalattuun aineistoon ja nopeusennusteet normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon.

9 Ruuhkan ennustaminen

Kuten aikaisemmissa luvuissa esitellyistä malleista on jo käynyt ilmi, etenkin nopeusennusteissa oli systemaattista virhettä ja malleilla oli taipumus yliarvioida alhaisia nopeuksia. Edellisessä kappaleessa esitetty naiivi malli ei jo määritelmänsä mukaan voi mukautua dynaamisiin, epäsäännöllisiin muutoksiin. Naiivin mallin tulokset olivat kuitenkin hyvin lähellä neuroverkkomalleilla saavutettuja tuloksia. Keskimääräiset virheet eivät kuitenkaan välttämättä kerro koko totuutta. Tästä syystä on mielenkiintoista nähdä, kuinka hyvin neuroverkkomallit suoriutuvat ruuhka-ajan liikennetilanteen ennustamisesta.

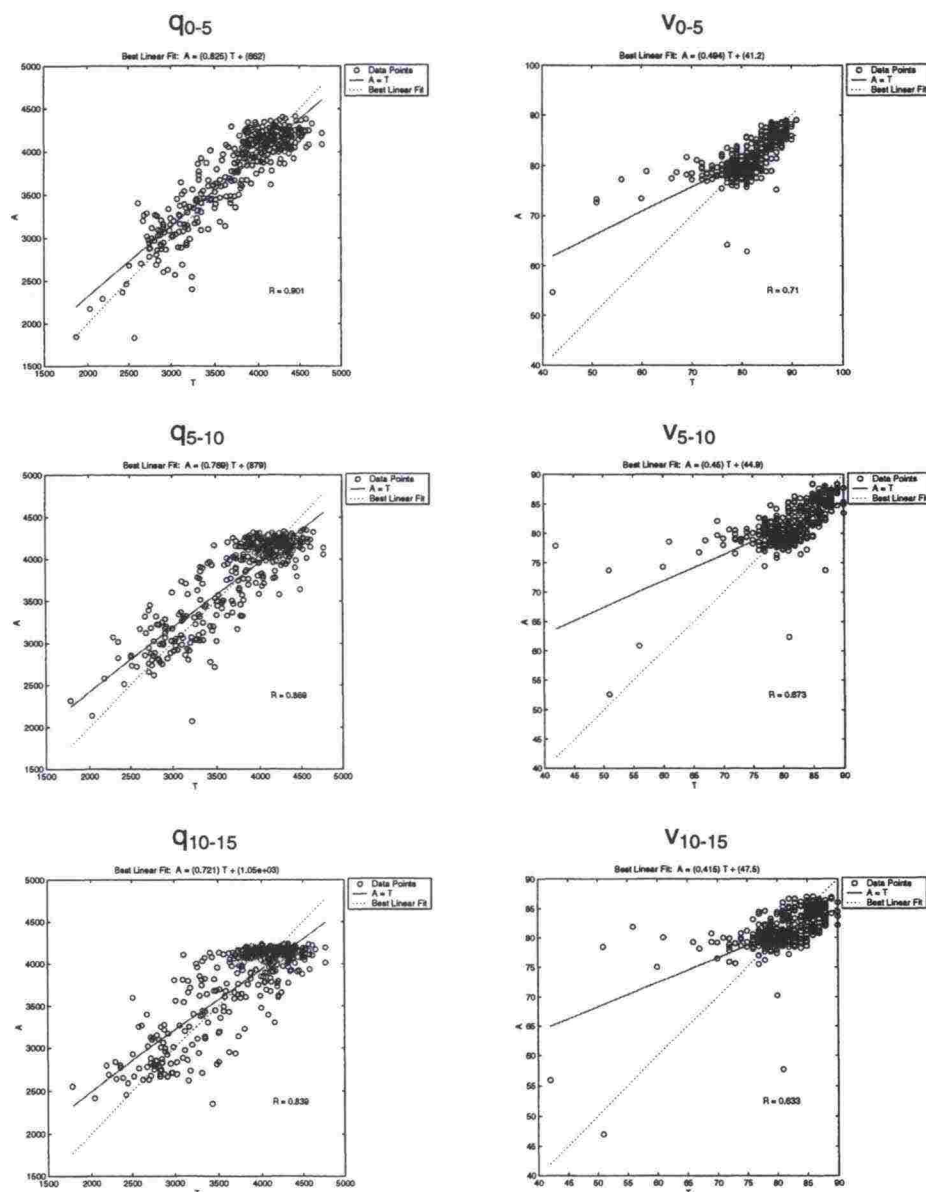
Ensimmäisenä ruuhkamallina kokeiltiin varta vasten ruuhka-aineistolla opetetua mallia. Ruuhka-ajaksi määriteltiin arkiaamujen klo 7 - 9 välinen liikenne. Malli oli keskihajontatietoja hyödyntävien mallien mukainen, eli sai liikennemäärä- ja keskinopeustietojen lisäksi ennustepoikkileikkauksen nopeuskeskihajontatiedot viimeiseltä 15 minuutilta. Malli jaettiin kahteen osamalliin, joista toinen ennusti liikennemäärää ja toinen nopeutta. Opetusaineiston pienen koon vuoksi (374 havaintoa) päätettiin kuitenkin tyytyä vain kahden poikkileikkauksen malliin. Lisätietopoikkileikkaukseksi valittiin parhaan tuloksen antanut poikkileikkaus 6. Ennuste tehtiin poikkileikkaukselle 11, kuten muissakin malleissa.

Taulukossa 43 on esitetty keskihajontatietoja hyödyntävän ruuhka-aineistolla opetetun mallin tulokset erilaisten virhetermien avulla esitettynä. Taulukosta nähdään, että skaalattuun aineistoon perustuva liikennemäärämalli oli parempi kuin normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuva malli. Nopeusennusteissa tilanne oli päinvastainen.

Taulukko 43. Poikkileikkausten 11 ja 6 aineistoon perustuvien ruuhka-aineistolla opetettujen ennustemallien virhetermit. Taulukossa normeeratulla aineistolla tarkoitetaan normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa.

Virhetermi	Aikajakso	Liikennemäärä		Nopeus	
		Skaalattu	Normeerattu	Skaalattu	Normeerattu
Keskimääräinen neliövirhe	0 - 5	59 000	57 000	14	16
	5 - 10	81 000	81 000	18	17
	10 - 15	107 000	111 000	18	18
Keskimääräinen absoluuttinen virhe	0 - 5	0	-17	1	0
	5 - 10	9	-5	1	0
	10 - 15	1	-8	0	0
Keskimääräinen absoluuttisen virheen itseisarvo	0 - 5	190	190	3	3
	5 - 10	220	230	3	3
	10 - 15	260	270	3	3
Keskimääräinen suhteellinen virhe	0 - 5	0 %	0 %	1 %	0 %
	5 - 10	1 %	1 %	1 %	1 %
	10 - 15	1 %	1 %	1 %	0 %
Keskimääräinen suhteellisen virheen itseisarvo	0 - 5	5 %	5 %	3 %	3 %
	5 - 10	6 %	6 %	4 %	4 %
	10 - 15	7 %	7 %	4 %	4 %

Skaalattuun aineistoon perustuvalle liikennemäärän ennustamismallille ja normeerattuun pääkomponenttianalyysin avulla käsiteltyyn aineistoon perustuvalle keskinopeuden ennustamismallille tehtiin regressiomalli (kuva 24). Regressioanalyysin kuvaajista näkee, kuinka ennuste huononi ennustejakson pidentessä.



Kuva 24. Regressioanalyysi nopeuskeskihajontatietoja hyödyntävälle ruuhka-aineistolla opetetulle mallille. X-akselilla on "oikea" ja y-akselilla ennustettu arvo. Liikennemäärää on merkitty q:lla ja keskinopeutta v:llä. Alaindeksit viittaavat ennustejaksoon.

Kiinnostavaa on myös nähdä, kuinka koko päivän aineistolla opetetut mallit suoriutuivat aamuruuhkan (arkipäivien klo 7 - 9 välinen liikenne) ennustamisesta. Ennusteet tehtiin nopeuskeskihajontatietoja hyödyntävällä mallilla. Taulukossa 44 on esitetty mallin tekemien ruuhka-ajan ennusteiden virhetermit ja saman mallin koko päivän liikenteen ennusteista lasketut virhetermit. Taulukosta havaitaan, että ruuhka-ajan ennusteissa keskimääräiset absoluuttiset virheet olivat suurempia kuin keskimäärin. Mallilla oli taipumus aliarvioida sekä liikennemäärä-että keskinopeusennusteita. Liikennemäärän suhteelliset virheet olivat pienempiä ruuhka-aikaan kuin koko päivän vastaava luku. Tämä oli seurausta ruuhka-ajan suurista liikennemääristä. Nopeudessa sitä vastoin ruuhka-ajan keskimää-

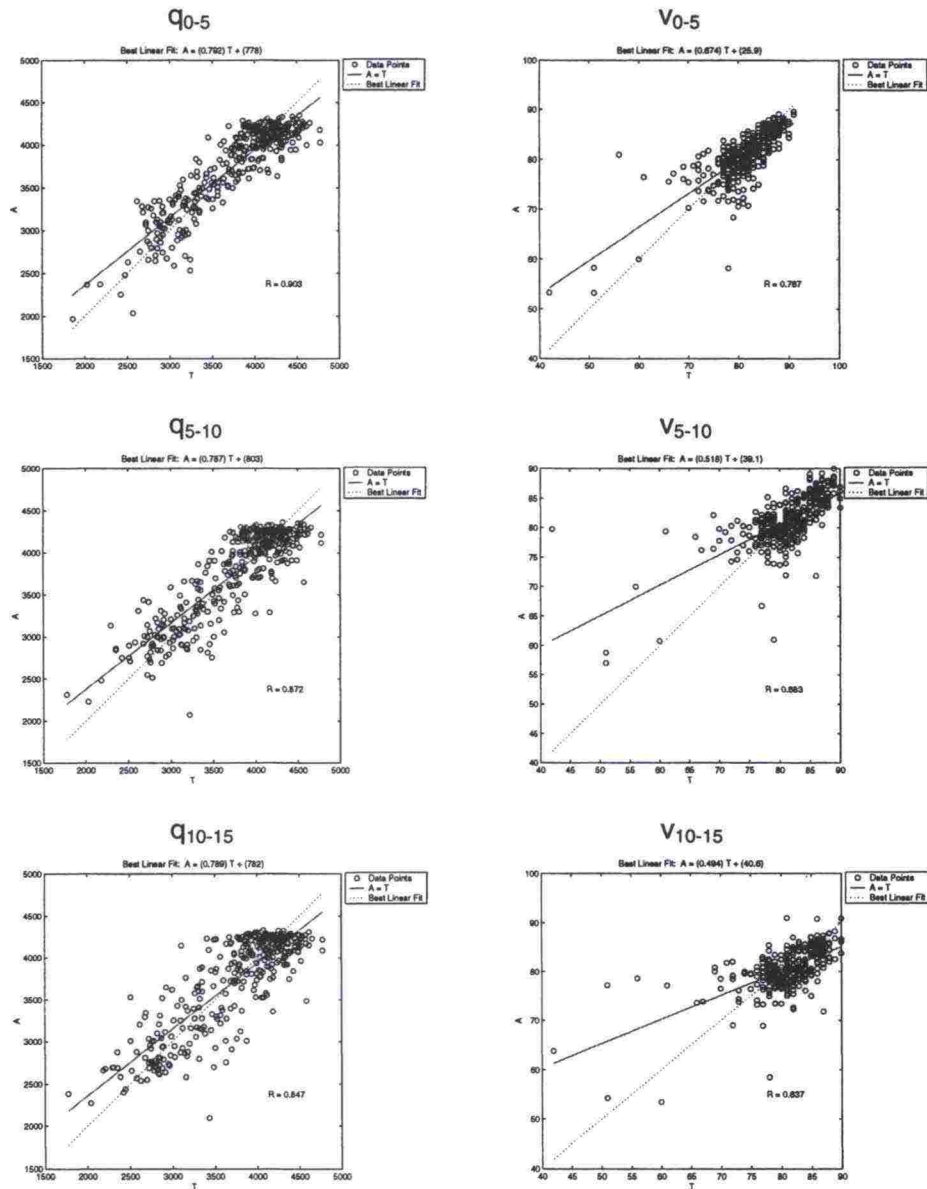
räinen suhteellisen virheen itseisarvo oli suurempi kuin koko päivän vastaava luku.

Taulukko 44. Nopeuskeskihajontatietoja hyödyntävän mallin tekemät ruuhka-ajan liikenteen ennusteen virhetermit sekä saman mallin koko päivän liikenteen ennusteista lasketut keskimääräiset virhetermit.

Virhetermi	Aikajakso	Liikennemäärä		Nopeus	
		Ruuhka	Koko päivä	Ruuhka	Koko päivä
Keskimääräinen neliövirhe	0 - 5	58 000	36 000	14	5
	5 - 10	80 000	43 000	17	6
	10 - 15	105 000	52 000	19	7
Keskimääräinen absoluuttinen virhe	0 - 5	-11	1	-1	0
	5 - 10	-1	-3	0	0
	10 - 15	-12	3	-1	0
Keskimääräinen absoluuttisen virheen itseisarvo	0 - 5	190	150	3	2
	5 - 10	220	160	3	2
	10 - 15	250	180	3	2
Keskimääräinen suhteellinen virhe	0 - 5	0 %	2 %	-1 %	0 %
	5 - 10	1 %	2 %	0 %	0 %
	10 - 15	0 %	2 %	0 %	0 %
Keskimääräinen suhteellisen virheen itseisarvo	0 - 5	5 %	9 %	3 %	2 %
	5 - 10	6 %	10 %	4 %	2 %
	10 - 15	7 %	11 %	4 %	2 %

Kun nopeuskeskihajontamallin ruuhka-ajan ennusteita (taulukko 44) verrataan varsinaisen ruuhkamallin tekemiin ennusteisiin (taulukko 43), havaitaan, että koko päivän aineistolla opetettu malli teki keskimääräisen neliövirheen perusteella mitattuna hieman parempia liikennemääräennusteita kuin ruuhkamalli. Keskimääräisellä absoluuttisella virheellä mitattuna tilanne oli kuitenkin päinvastainen. Muissa virhetermeissä erot olivat liikennemäärän osalta pieniä. Nopeusennusteista molemmat mallit suoriutuivat suunnilleen yhtä huonosti.

Kuvassa 25 on esitetty nopeuskeskihajontatietoja hyödyntävän koko päivän liikenteeseen perustuvan mallin ruuhkaennusteille regressioanalyysi. Kun näitä tuloksia verrataan ruuhka-aineistolla opetetun mallin tuloksiin (kuva 24), havaitaan, että molemmat mallit suoriutuivat liikennemääräennusteista suunnilleen yhtä hyvin. Liikennemääräennuste oli ensimmäiselle ennustejaksolle tarkin ja viimeiselle huonoin. Nopeusennusteista sitä vastoin koko päivän aineistolla opetettu malli tuntui suoriutuvan - etenkin alhaisimpien nopeuksien osalta - paremmin kuin pelkällä ruuhka-aineistolla opetettu malli. Myös nopeusennusteissa ennustejaksoista ensimmäisellä ennuste oli tarkin ja viimeisellä huonoin.

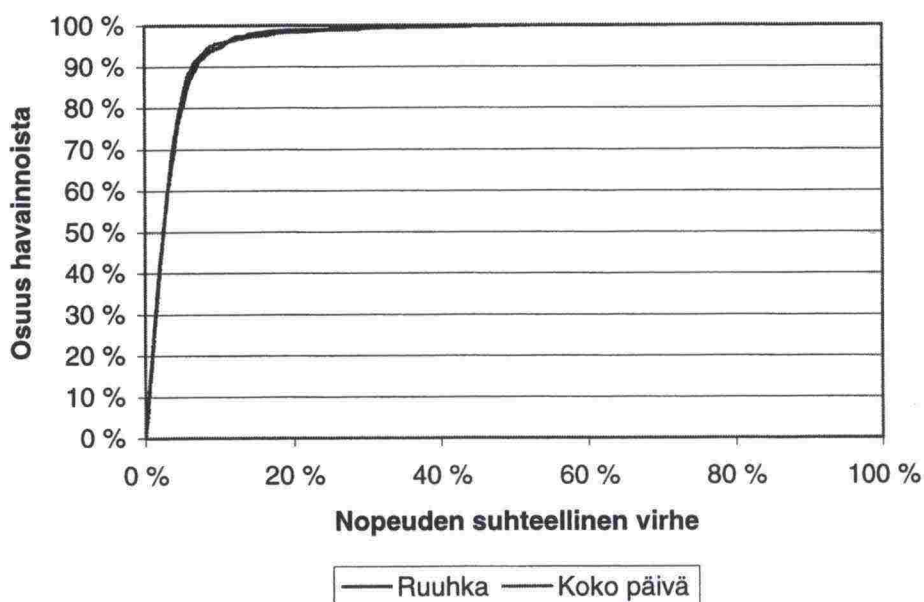
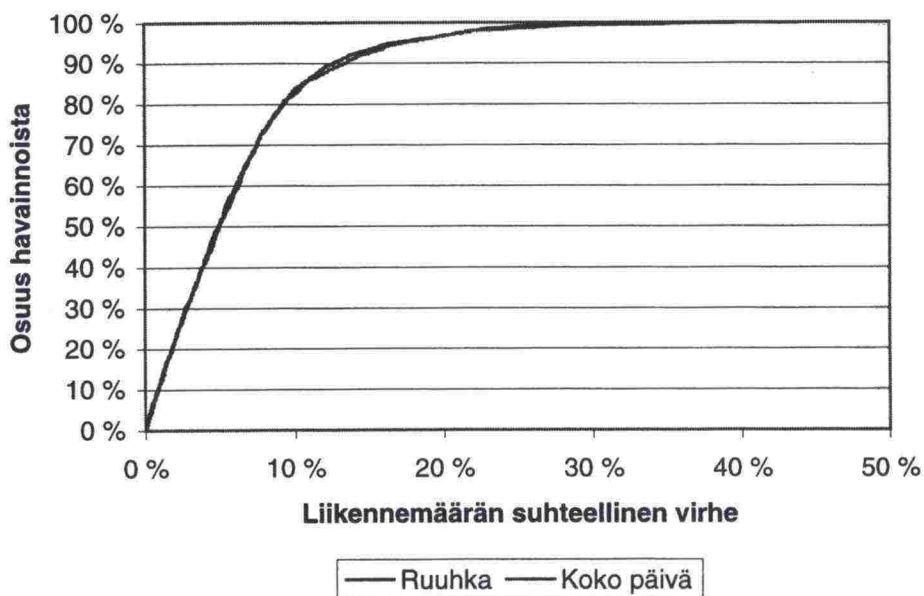


Kuva 25.

Regressioanalyysi nopeuskeskihajontatietoja hyödyntävän koko päivän liikenteeseen perustuvan mallin tekemille ruuhkaliikenteen ennusteille. X-akselilla on "oikea" ja y-akselilla ennustettu arvo. Liikennemäärää on merkitty q :lla ja keskinopeutta v :llä. Alaindeksit viittaavat ennustejaksoon.

Kuvassa 26 on esitetty koko päivän liikenteeseen ja ruuhka-ajan liikenteeseen perustuvien mallien suhteellisten virheiden itseisarvojen kumulatiiviset summakäyrät aamuruuhkan ennusteille. Koko päivän aineistoon perustuvalla mallilla 90 prosentissa liikennemääräennusteissa suhteellinen virhe oli korkeintaan 12 prosenttia ja ruuhka-ajan aineistoon perustuvalla mallilla 13 prosenttia. Nopeusennusteissa koko päivän aineistoon perustuvalla mallilla vastaava luku oli seitsemän prosenttia ja ruuhka-ajan aineistoon perustuvalla mallilla kuusi prosenttia. Suhteellinen virhe oli korkeintaan kymmenen prosenttia 84 prosentissa ruuhka-aineistoon perustuvan mallin liikennemääräennusteista ja 85 prosenttia koko

päivän aineistoon perustuvan mallin ennusteista. Vastaava osuus ruuhka-ajan aineistoon perustuvan mallin nopeusennusteista oli 96 prosenttia ja koko päivän aineistoon perustuvan mallin nopeusennusteista 95 prosenttia.



Kuva 26.

Aamuruuhkan liikennemäärän ja keskinopeuden suhteellisten virheiden itseisarvojen kumulatiiviset summakäyrät nopeuskeskihajontatietoja hyödyntävälle ruuhkamallille ja vastaavalle koko päivän liikennetietoihin perustuvalle mallille.

10 Sovellus kaksikaistaiselle tielle

10.1 Yleistä

Edellä saatuja tuloksia sovellettiin kaksikaistaiselle tielle. Kohteena oli valtatie 3 Valkeakoskella (*kuva 2*). Ennustepisteestä (LAM-piste 433, suuntaan 1 Tampereelle) lisätietopisteeseen (LAM-piste 402) oli matkaa 17,6 kilometriä.

LAM-pisteiden välillä olevista liittymistä lähempänä LAM-pistettä 433 oleva Valkeakoskelle menevän tien liittymä (*kuva 2*) on vilkasliikenteisin. Siellä keskimääräinen vuorokausiliikenne on 7 400 ajon./vrk. Samasta liittymästä Toijalaan menevällä tiellä vastaava liikennemäärä on 3 900 ajon./vrk. Valkeakosken ja Toijalan toisten liittymien liikennemäärät ovat 2 400 - 2 500 ajon./vrk. Sääksmäelle menevän tien liikennemäärä on pienempi. Tarkastelualueella valtatie 3 liikennemäärä on 13 000 ajon./vrk. (Tielaitos 1999.)

Valtatielle 3 tehtiin liikennemäärä- ja keskinopeusennusteet Länsiväylän mallin mukaisesti. Tämän lisäksi tehtiin malli, joka ennusti jonoprosenttia, koska myös jonossa ajavien ajoneuvojen osuus kuvaa liikenteen sujuvuutta kaksikaistaisella tiellä.

Malleille päätettiin antaa syötetietoina kustakin pisteestä puolen tunnin liikennemäärä- ja keskinopeustiedot. Perusteina pidempään aikasarjaan oli LAM-pisteiden Länsiväylän pisteisiin verrattuna pitkä välimatka ja toisaalta käytettävissä olevan aineiston suurempi määrä (yhteensä 36 858 havaintoa). Jonoprosenttimalli sai lisätietoina ennustesuunnan LAM-pisteiden viiden minuutin keskimääräiset jonoprosentit samalta puolen tunnin jaksolta.

Mallit tehtiin neljään erilaiseen aineistoon perustuviksi. Ensimmäisessä aineistossa oli liikennetiedot LAM-pisteen 433 molemmilta suunnilta sekä lisänä vastaavat tiedot LAM-pisteen 402 suunnalta 1 (ennustesuunta). Toisessa aineistossa oli molempien LAM-pisteiden liikennetiedot suunnalta 1. Kolmannessa aineistossa oli LAM-pisteen 433 liikennetiedot molemmilta suunnilta. Viimeisessä aineistossa oli ainoastaan LAM-pisteen 433 suunnan 1 liikennetiedot.

Malleissa päätettiin käyttää skaalattua aineistoa. Ensimmäisessä liikennemäärän ja keskinopeuden ennustamismallissa oli 36 syöte- ja 42 piiloneuronia. Toisessa ja kolmannessa mallissa syöteneuroneja oli 24 ja piiloneuroneja 59. Viimeisessä mallissa syöteneuroneja oli 12 ja piiloneuroneja 96. Kaikissa em. malleissa oli kuusi vastaneuronia.

Aineistoon 1 perustuvassa jonoprosentin ennustamismallissa oli 48 syöte- ja 35 piiloneuronia. Aineistoon 2 perustuvassa mallissa syöteneuroneja oli 36 ja piiloneuroneja 45. Aineistoon 3 perustuvassa mallissa vastaavat luvut olivat 30 ja 54 sekä aineistoon 4 perustuvassa mallissa 18 ja 83. Kaikissa em. malleissa oli kolme vastaneuronia.

10.2 Opetusalgorithmi

Mallien opettamiseen käytettiin Fletcher-Reevesin päivitystä, joka kuuluu konjugaatti-gradienttimenetelmiin. Syynä opetusalgoritmin vaihtoon oli se, että edellä käytetty Levenberg-Marquardt-algoritmi on nopea ainoastaan muutamia satoja painokertoimia sisältävien neuroverkkojen opetuksessa ja nyt tehdyissä malleissa painokertoimia oli noin 1 800. Näin suurella neuroverkolla laskenta vaatisi Levenberg-Marquardt-algoritmeilla valtavasti muistia (Demuth & Beale 1998).

Perusgradienttimenetelmissä neuroverkon painokertoimia päivitetään siihen suuntaan, jossa hyvyysfunktio (tässä tapauksessa virheen neliösumma) pienee nopeimmin. Tämä ei kuitenkaan välttämättä takaa nopeinta konvergenssia. Konjugaatti-gradienttimenelmässä suunnaksi valitaan konjugaatin suunta, jolloin päästään nopeampaan konvergenssiin kuin suurimman gradientin suunnassa. (Demuth & Beale 1998.)

Kaikissa konjugaatti-gradienttimenetelmissä algoritmi aloittaa ensimmäisellä iteroitinkierroksella etsimisen nopeimman laskun suuntaan.

$$\mathbf{p}_0 = -\mathbf{g}_0$$

Tämän jälkeen määritellään optimaalinen etäisyys, jonka verran valittuun suuntaan siirrytään.

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k$$

Nyt seuraava etsimissuunta määritellään siten, että se on konjugaatiossa aiemman etsintäsuunnan kanssa. Uusi etsimissuunta löytyy siis yhdistämällä uusi nopeimman laskun suunta edelliseen etsimissuuntaan.

$$\mathbf{p}_k = -\mathbf{g}_k + \beta_k \mathbf{p}_{k-1}$$

Erilaiset konjugaatti-gradienttimenetelmien versiot eroavat toisistaan siinä, miten vakio β_k lasketaan. Fletcher-Reevesin päivityksessä tämä tapahtuu seuraavasti.

$$\beta_k = \frac{\mathbf{g}_k^T \mathbf{g}_k}{\mathbf{g}_{k-1}^T \mathbf{g}_{k-1}}$$

β_k on siis nykyisen gradientin neliönormin suhde aiemman gradientin neliönormiin. (Demuth & Beale 1998.)

10.3 Liikennemäärä- ja nopeusennusteiden tulokset

Taulukoissa 45 ja 46 on esitetty liikennemäärän ja keskinopeuden ennustamismallien tulokset eri virhetermien avulla. Tulokset pisteytettiin vastaavasti kuin edellä siten, että parhaasta tuloksesta (= pienin virhetermi) sai 3 pistettä, toiseksi parhaasta 2 pistettä ja niin edelleen. Nämä mallin hyvyyspisteet on annettu taulukossa 47. Taulukosta havaitaan, että toisen mittauspisteen liikennetiedot

olivat tärkeämmät kuin ennustepisteen vastakkaisen suunnan liikennetiedot. Paras tulos saatiinkin, jos käytettävissä oli molempien mittauspisteiden liikennetiedot ennustesuunnalta. Lähes yhtä hyvään tulokseen ylsi malli, joka sai em. tietojen lisäksi liikennetiedot ennustepisteestä ennustesuuntaa vastaan ajavasta liikenteestä. Tulokset olivat heikoimmat, kun syötetietoa oli vähiten (ainoastaan ennustepisteen liikennetiedot ennustesuunnalta).

Taulukko 45. Virhetermit eri aineistoihin perustuvilla liikennemäärän ennustamismalleille, jotka on rakennettu valtatie 3 LAM-pisteen 433 suunnalle 1. Aineistossa 1 on liikennetiedot molemmilta suunnilta LAM 433:sta ja suunnalta 1 LAM 402:sta, aineistossa 2 on liikennetiedot molemmista pisteistä suunnalta 1, aineistossa 3 molempien suuntien ja aineistossa 4 ainoastaan suunnan 1 liikennetiedot LAM 433:sta.

Virhetermi	Aikajakso	Liikennemäärä			
		Aineisto 1	Aineisto 2	Aineisto 3	Aineisto 4
Keskimääräinen neliövirhe	0 – 5	6 200	6 200	7 800	8 100
	5 – 10	6 200	6 300	8 000	8 500
	10 – 15	7 000	7 100	8 500	9 100
Keskimääräinen absoluuttinen virhe	0 – 5	0	0	0	-1
	5 – 10	-1	0	0	0
	10 – 15	0	0	0	-1
Keskimääräinen absoluuttisen virheen itseisarvo	0 – 5	59	59	66	67
	5 – 10	58	59	67	69
	10 – 15	63	63	69	71
Keskimääräinen suhteellinen virhe	0 – 5	10 %	9 %	12 %	11 %
	5 – 10	10 %	9 %	12 %	13 %
	10 – 15	11 %	11 %	13 %	13 %
Keskimääräinen suhteellisen virheen itseisarvo	0 – 5	25 %	24 %	28 %	29 %
	5 – 10	24 %	24 %	28 %	30 %
	10 – 15	27 %	27 %	29 %	31 %

Taulukko 46. Virhetermit eri aineistoihin perustuvilla keskinopeuden ennustamismalleille, jotka on rakennettu valtatie 3 LAM-pisteen 433 suunnalle 1. Aineistossa 1 on liikennetiedot molemmilta suunnilta LAM 433:sta ja suunnalta 1 LAM 402:sta, aineistossa 2 on liikennetiedot molemmista pisteistä suunnalta 1, aineistossa 3 molempien suuntien ja aineistossa 4 ainoastaan suunnan 1 liikennetiedot LAM 433:sta.

Virhetermi	Aikajakso	Nopeus			
		Aineisto 1	Aineisto 2	Aineisto 3	Aineisto 4
Keskimääräinen ne- liövirhe	0 – 5	20	20	21	22
	5 – 10	21	21	22	23
	10 – 15	22	22	22	23
Keskimääräinen ab- soluuttinen virhe	0 – 5	0	0	0	0
	5 – 10	0	0	0	0
	10 – 15	0	0	0	0
Keskimääräinen ab- soluuttisen virheen itseisarvo	0 – 5	3	3	3	3
	5 – 10	3	3	3	3
	10 – 15	3	3	3	3
Keskimääräinen suhteellinen virhe	0 – 5	0 %	0 %	0 %	0 %
	5 – 10	0 %	0 %	0 %	0 %
	10 – 15	0 %	0 %	0 %	0 %
Keskimääräinen suhteellisen virheen itseisarvo	0 – 5	4 %	4 %	4 %	4 %
	5 – 10	4 %	4 %	4 %	4 %
	10 – 15	4 %	4 %	4 %	4 %

Taulukko 47. Taulukon 46 tuloksista annetut hyvyyspisteet kullekin aineistolle.

Aineisto	Hyvyyspisteet
Aineisto 1 LAM-piste 433: molemmat suunnat, LAM-piste 402: suunta 1	58
Aineisto 2 Molemmat LAM-pisteet: suunta 1	74
Aineisto 3 LAM-piste 433: molemmat suunnat	37
Aineisto 4 LAM-piste 433: suunta 1	11

Koska mallin opetusdatassa oli mukana myös yöajan liikenne, jossa liikennemäärät ovat erittäin pieniä, suhteellinen virhe ei kerro koko totuutta. Pienellä liikennemäärällä pienikin absoluuttinen virhe voi olla suhteellisenä virheenä ilmaistuna hyvin suuri, joten suhteellisten virheiden sijaan kumulatiivisten summakäyrien avulla tarkasteltiin absoluuttisia virheitä (taulukko 48).

Taulukko 48. Valtatieltä 3 koottuihin eri aineistoihin perustuvien mallien absoluuttisen virheen itseisarvojen tunnuslukuja. Aineistossa 1 on liikennetiedot molemmilta suunnilta LAM 433:sta ja suunnalta 1 LAM 402:sta, aineistossa 2 on liikennetiedot molemmista pisteistä suunnalta 1, aineistossa 3 molempien suuntien ja aineistossa 4 ainoastaan suunnan 1 liikennetiedot LAM 433:sta.

		Aineisto 1	Aineisto 2	Aineisto 3	Aineisto 4
Osuus q-ennusteista, joissa virhe korkeintaan...	100 ajon./h	82 %	82 %	77 %	76 %
	50 ajon./h	54 %	54 %	50 %	48 %
Osuus v-ennusteista, joissa virhe korkeintaan...	5 km/h	83 %	84 %	83 %	82 %
	1 km/h	32 %	32 %	31 %	30 %
Virhe, jonka 10 % ennusteista ylittää	q (ajon./h)	129	129	144	147
	v (km/h)	7	7	7	7

Aineistoon 2 perustuvan mallin tulokset olivat parhaat ja toiseksi ylsi aineistoon 1 perustuva malli - ero tosin oli erittäin pieni. Näistä selvästi heikommin suoriutuvat aineistoon 3 ja 4 perustuvat mallit, joista jälkimmäinen oli hieman ensin mainittua huonompi. Myös tästä voidaan tehdä johtopäätös, että oli tärkeämpää saada yhden suunnan liikennetiedot kahdesta pisteestä kuin yhdestä pisteestä molemmilta ajosuunnilta.

10.4 Jonoprosenttiennusteiden tulokset

Taulukossa 49 on esitetty eri aineistoihin perustuvien jonoprosentin ennustamismallien tulokset. Havaitaan, että myös jonoprosentin ennustamisessa aineistoihin 1 ja 2 perustuvat mallit olivat parempia kuin aineistoihin 3 ja 4 perustuvat mallit. Kaikkien mallien keskimääräinen absoluuttinen virhe oli nolla prosenttia eli mallit eivät keskimääräin yli- tai aliarvioineet ennusteita systemaattisesti. Keskimääräinen absoluuttisen virheen itseisarvo oli aineistoon 1 ja 2 perustuvilla malleilla kahdeksan - yhdeksän prosenttia ja aineistoihin 3 ja 4 perustuvilla malleilla yhdeksän prosenttia.

Taulukko 49. Virhetermit eri aineistoihin perustuville jonoprosentin ennustamismalleille, jotka on rakennettu valtatie 3 LAM-pisteen 433 suunnalle 1. Aineistossa 1 on liikennetiedot molemmilta suunnilta LAM 433:sta ja suunnalta 1 LAM 402:sta, aineistossa 2 on liikennetiedot molemmista pisteistä suunnalta 1, aineistossa 3 molempien suuntien ja aineistossa 4 ainoastaan suunnan 1 liikennetiedot LAM 433:sta.

Virhetermi	Aikajakso	Jonoprosentti			
		Aineisto 1	Aineisto 2	Aineisto 3	Aineisto 4
Keskimääräinen neliövirhe	0 – 5	0,011	0,011	0,013	0,013
	5 – 10	0,011	0,011	0,013	0,013
	10 – 15	0,012	0,012	0,013	0,013
Keskimääräinen absoluuttinen virhe	0 – 5	0 %	0 %	0 %	0 %
	5 – 10	0 %	0 %	0 %	0 %
	10 – 15	0 %	0 %	0 %	0 %
Keskimääräinen absoluuttisen virheen itseisarvo	0 – 5	8 %	8 %	9 %	9 %
	5 – 10	8 %	8 %	9 %	9 %
	10 – 15	9 %	9 %	9 %	9 %

Eri aineistoihin perustuvien mallien tuloksia vertailtiin myös absoluuttisen virheen itseisarvojen kumulatiivisten summakäyrien avulla (taulukko 50). 90 prosentissa aineistoihin 1 ja 2 perustuvien mallien ennusteista virhe oli korkeintaan 17 prosenttia ja aineistoihin 3 ja 4 perustuvien mallien ennusteista 18 prosenttia. Aineistoon 1 perustuvan mallin ennusteista 69 prosentissa virhe oli korkeintaan 10 prosenttia. Aineistoon 2 perustuvalla mallilla vastaava osuus oli 68 prosenttia, aineistoon 3 perustuvalla mallilla 66 prosenttia ja aineistoon 4 perustuvalla mallilla 65 prosenttia.

Taulukko 50. Eri aineistoihin perustuvien jonoprosentin ennustamismallien absoluuttisen virheen itseisarvojen tunnuslukuja. Aineistossa 1 on liikennetiedot molemmilta suunnilta LAM 433:sta ja suunnalta 1 LAM 402:sta, aineistossa 2 on liikennetiedot molemmista pisteistä suunnalta 1, aineistossa 3 molempien suuntien ja aineistossa 4 ainoastaan suunnan 1 liikennetiedot LAM 433:sta.

	Aineisto 1	Aineisto 2	Aineisto 3	Aineisto 4
Osuus ennusteista, joissa virhe korkeintaan 10 %	69 %	68 %	66 %	65 %
Virhe, jonka 10 % ennusteista ylittää	17 %	17 %	18 %	18 %

Yhteenvedona voidaan siis todeta, että myös jonoprosentin ennustamisessa oli huomattavasti tärkeämpää saada liikennetiedot useammasta poikkileikkauksesta ennustesuunnalta kuin saada vastakkaisen suunnan liikennetiedot. Vastakkaisen suunnan liikennetiedot paransivat mallin suoritusta hieman, mutta ero oli pieni.

Jos kaksikaistaiselle tielle siis halutaan tehdä ennustemalli, kannattaa ilmaisimet sijoittaa mieluummin ennustesuunnalle useammalle poikkileikkaukselle kuin kattaa molemmat suunnat. Vastaantulevan liikenteen tietojen vaikutus mallin

suoritukseen oli suhteellisen pieni. Erityisen hyvin tämän voi havaita liikennemäärä- ja jonoprosenttiennusteissa, nopeusennusteissa erot olivat pienemmät.

11 Johtopäätöksiä

Tutkimuksessa haluttiin selvittää, minkälainen on lyhyen aikavälin liikennetilanne-ennusteen kannalta optimaalinen neuroverkko. Tämän jälkeen optimoitiin muita ennustemalliin liittyviä tekijöitä, kuten syötteitä, ennustejakson pituutta ja poikkileikkausten sijaintia. Lopuksi neuroverkkoon perustuvia malleja verrattiin naiiviin malliin ja Länsiväylällä kerättyjä kokemuksia sovellettiin kaksikaistaisessa kohteessa.

Ensimmäisenä tutkimushypoteesina oli, että syötedatan dimensioiden vähentäminen pääkomponenttianalyysin avulla kannattaa silloin, kun syöteparametrejä on paljon. Tämä ei kuitenkaan välttämättä pitänyt paikkaansa, sillä skaalaamalla päästiin monta kertaa jopa parempiin tuloksiin kuin käsittelemällä normeerattu aineisto pääkomponenttianalyysin avulla. Näin siitä huolimatta, että skaalattua aineistoa käyttävissä malleissa piiloneuroneiden määrä jäikin paljon pienemmäksi kuin pääkomponenttianalyysin avulla dimensioiltaan pienempää aineistoa käyttävillä malleilla.

Toisena hypoteesina oli, ettei aktivaatiofunktioyhdistelmien välillä ole eroa ennusteen tarkkuuden kannalta. Hypoteesi piti kuitenkin hylätä. Sekä skaalattua että normeerattua pääkomponenttianalyysin avulla käsiteltyä aineistoa käyttävillä malleilla parhaaksi aktivaatiofunktioyhdistelmäksi osoittautui piilokerroksen hyperbolinen tangentti yhdistettynä vastekerroksen lineaariseen funktioon.

Kolmantena optimaaliseen neuroverkkoon liittyvänä hypoteesina oli, että osamalleihin jaetun ennustemallin tulokset ovat vähintään yhtä hyviä kuin jakamattoman mallin. Tämä tulos voitiin vahvistaa sellaisen mallin avulla, jossa liikennemäärä- ja nopeusennusteet tehtiin eri osamalleissa. Mallin, jossa jokaiselle ennustejaksolle oli oma osamalli, tulokset olivat sitä vastoin huonommat kuin jakamattoman mallin. Tähän oli kuitenkin todennäköisesti ollut syynä opetuksen satunnaisuudesta seurannut sattuma. Joka tapauksessa erot jaettujen ja jakamattomien mallien välillä olivat hyvin pienet. Näin ollen jakamisesta aiheutuvan vaivan takia (yhden neuroverkon sijasta on opetettava kaksi tai kolme verkkoa) osamalleihin jakamisen avulla saatavaa hyötyä voidaan pitää kyseenalaisena.

Edellä mainitun tyypisiltä sattumilta (neuroverkolla ei päästy yhtä hyviin tuloksiin kuin teoriassa olisi pitänyt päästä) voidaan yrittää välttyä siten, että neuroverkko opetetaan useaan kertaan ja valitaan tulokseksi saaduista verkoista paras (kymmenen kertaa ei ilmeisesti ole riittävästi). Opetuksen onnistumista kannattaa myös tarkastella validointi- ja testidatan avulla. (Opetus voidaan katsoa epäonnistuneeksi, jos validointi- tai testidatan virhe on selvästi suurempi kuin opetusdatan.) Näin voidaan varmistua, että neuroverkkoa voidaan soveltaa myös opetusdatan ulkopuolella ilman, että tulokset kärsivät.

Hypoteesina oli, että kellonaika- ja päivätiedot parantavat ennusteita ja ettei sillä, miten em. tiedot annetaan, ole merkitystä ennusteen laadun kannalta. Saadut tulokset kuitenkin osoittivat, etteivät lisätiedot parantaneet ennusteen laatua.

Kun ottaa huomioon sen, että lisätiedot merkitsevät lisää neuroneja syötekerrokseen, jolloin piilokerroksen neuroneja täytyy vähentää, lisätiedot johtavat itse asiassa huonompiin tuloksiin kuin perusmalli, jossa piiloneuroneja oli enemmän. Suuremmalla opetusdatalla tilanne olisi tietenkin saattanut olla toinen, kun piiloneuroneja olisi lähtökohtaisesti ollut enemmän eikä muutaman piiloneuronin vähentämisen vaikutus olisi ollut yhtä suuri. Kellonaika- ja viikonpäivätiedon antotavalla ei hypoteesin mukaisesti näyttänyt olevan vaikutusta tuloksiin.

Syötetietoihin liittyvänä hypoteesina oli myös, että nopeuskeskihajontatiedot parantavat etenkin nopeusennusteita. Kun tarkastellaan regressioanalyysin tuloksia, voitiin hypoteesin mukaisesti havaita, että alhaisimpien nopeuksien osalta mallin antamat ennusteet olivat parempia kuin vertailumallin. Keskimääräisissä virhetermeissä eroa ei kuitenkaan huomattu.

Ennustejakson pituuteen liittyvänä hypoteesina oli, että ennusteen tarkkuus huononee ennustejakson pituuden kasvaessa. Näin ei kuitenkaan välttämättä ole; eräät virhetermit jopa pienenivät ennustejakson ajankohdan edetessä kauemmaksi tarkasteluhetkestä. Keskimäärin tarkasteltuna 30 minuutin ennuste oli hieman epätarkempi kuin 15 minuutin vertailuennuste, mutta esimerkiksi suhteellisten virheiden itseisarvojen summakäyriä tarkastellessa ero oli häviävän pieni.

Hypoteesina oli, että mitä kauempana lisätietopoikkileikkaus sijaitsee ennustepoikkileikkauksesta, sitä tarkempia ovat ennusteet. Kun poikkileikkauksia on kaksi ja lisätietopoikkileikkaus sijaitsi ennen ennustepoikkileikkausta (ennustepoikkileikkauksesta ylävirtaan), tämä piti paikkansa. Tilanne oli kuitenkin päinvastainen tilanteissa, joissa lisätietopoikkileikkaus sijaitsi ennustepoikkileikkauksen jälkeen.

Jos mallille halutaan antaa tiedot kolmelta poikkileikkaukselta ja ensimmäinen lisätietopoikkileikkaus on sijoitettu mahdollisimman kauas (kun puhutaan muutamien kilometrin välimatkasta) ennustepoikkileikkauksen taakse (ylävirtaan), toinen lisätietopoikkileikkaus kannatti sijoittaa ennustepoikkileikkauksesta hieman alavirtaan tai vaihtoehtoisesti ennustepoikkileikkauksen ja ensimmäisen lisätietopoikkileikkauksen puoliväliin.

Toisena hypoteesina poikkileikkausten sijoittamisessa oli, että on olemassa jokin etäisyys, jota lähemmäksi poikkileikkauksia ei kannata sijoittaa silloin, kun niiden määrä on rajallinen. Kun poikkileikkauksia oli kaksi ja lisätietopoikkileikkaus sijaitsi ennen ennustepoikkileikkausta (siitä ylävirtaan), lisätietopoikkileikkauksen minimietäisyydeksi saatiin tarkastellussa kohteessa suunnilleen 800 metriä. Maksimietäisyyttä ei käytettävissä olevan aineiston avulla voitu selvittää.

On kuitenkin hyvä pitää mielessä, että poikkileikkausten optimaalinen sijainti on voimakkaasti riippuvainen kulloisestakin kohteesta ja ennustejakson pituudesta, eli ettei nyt saatuja tuloksia voi välttämättä soveltaa suoraan muualla. Yleisesti

voidaan kuitenkin sanoa, että on olemassa paikallisista oloista riippuvaiset optimaaliset sijainnit, joihin lisätietopoikkileikkaukset kannattaa sijoittaa.

Neuroverkkomalleille tehtiin vertailumalliksi ns. naiivi malli. Malli perustui liikenteen säännönmukaisuuteen ja sen tekemä ennuste pohjautui kunakin päivätyyppinä kullakin ajanhetkellä keskimääräisiin liikennemääriin ja keskinopeuksiin. Naiiviin malliin liittyvinä hypoteeseina oli, että neuroverkkomallin antamat ennusteet ovat tarkempia kuin naiivin mallin, koska naiivi malli ei pysty reagoimaan odottamattomiin liikenteen häiriöihin. Tämän näki erityisen selvästi nopeusennusteista, joiden arvo ei naiivissa mallissa koskaan laskenut alle tietyn raja-arvon (noin 74 km/h). Naiivi malli ei siis koskaan voi ennustaa alhaisia ruuhkanopeuksia, jollei ruuhka ole toistuvaa.

Liikennemääräennusteissa naiivi malli sitä vastoin onnistui paljon paremmin kuin nopeusennusteissa. Vaikka naiivi malli epäonnistuiakin säännöllisesti häiriö- ja ruuhkatilanteiden ennustamisessa, näiden osuus havainnoista oli niin pieni, että naiivin mallin ja sen vertailumallin välinen ero keskimääräisissä virhetermeissä jäi pieneksi. Olennainen ero mallien välillä on kuitenkin se, että riittävällä aineistolla ja oikeilla syötteillä neuroverkkomalli on mahdollista opettaa reagoimaan satunnaisiin häiriötilanteisiin, mutta naiivissa mallissa tämä ei ole edes teoriassa mahdollista.

Hypoteesina oli, että ruuhka-ajan aineistolla opetettu malli tekee parempia ruuhka-ajan ennusteita kuin koko päivän aineistolla opetettu malli. Hypoteesi piti kuitenkin hylätä, koska saadut tulokset olivat yhtä hyviä. Tilanne olisi saattanut olla toinen, jos opetusjoukko olisi ollut suurempi. Nyt käytetyssä aineistossa oli ainoastaan 374 havaintoa arki-aamujen ruuhkasta.

Kaksikaistaiselle tielle tehdyn ennustemallin perusteella havaittiin, että on tärkeämpää saada tarkastellun ajosuunnan liikennetiedot kahdesta pisteestä kuin yhdestä pisteestä molemmilta suunnilta. Vastaantulevan liikenteen tietojen vaikutus mallin suoritukseen oli suhteellisen pieni. Tähän oli ilmeisesti syynä se, että liikennemäärät pysyivät pääosin varsin pieninä suhteessa tien välityskykyyn eikä vastaantuleva liikenne tällöin juurikaan vaikuttanut liikennetilanteeseen. Tämän pienen kokeilun perusteella vaikuttaisi siltä, että Länsiväylältä kerättyjä kokemuksia voidaan soveltaa suurelta osin suoraan myös kaksikaistaisilla teillä.

Ennustemallia suunniteltaessa mallille ei kannata antaa syöteparametreinä kaikkea mahdollista tietoa. Etenkin silloin, jos opetusdatan keräysjakso on lyhyt, kannattaa poikkileikkausten määrä ja mukaan otettavat parametrit miettiä tarkkaan. Nyt käytetyllä aineistolla kahden poikkileikkauksen liikennetietoja käytettävällä neuroverkolla päästiin hyvin lähelle kaikkien seitsemän poikkileikkauksen liikennetietoja käyttävää mallia ja viikonpäivä- ja kellonaikatietoja hyväksi käyttävien mallien tulokset jäivät huonommiksi kuin perusmalleilla.

Sää ja keli vaikuttavat voimakkaasti Länsiväylän liikenteen käyttäytymiseen (Innamaa ym. 2000). Nyt koottu aineisto oli säiden suhteen melko homogeeniselta

jaksolta (toukokuu). Jos ennustemallia halutaan käyttää ympäri vuoden, sen on toimittava kaikenlaisissa olosuhteissa. Tällöin myös opetusdata on kerättävä ajanjaksolta, johon mahtuu kaikenlaisia säitä ja kelejä. Tällaisella aineistolla opetettaessa sää- ja kelityypistä kannattaa tehdä neuroverkkomallille oma syöteparametri.

Levenberg-Marquardt:n algoritmilla opetettaessa opetuskierrokset jäivät usein alle 20:een ja enimmilläänkin niitä oli 100 – 200. Opetus päättyi lähestulkoon aina validointidatan virheen kasvuun. Esimerkiksi Zhang ym. (1997) opettivat omaa neuroverkkoaan 10 miljoonan opetuskierroksen verran. Vaikka tarvittavien opetuskierrosten lukumäärä onkin voimakkaasti riippuvainen opetusjoukon koosta, käytetystä opetusmenetelmästä ja sen tehokkuudesta, luku tuntuu erittäin suurelta. Näin monen opetuskierroksen jälkeen on riskinä, että neuroverkko on oppinut datajoukon ulkoa, jolloin tulokset kyllä ovat erinomaisia, mutta malli ei osaa yleistää opetusdatan ulkopuolelle.

Neuroverkkojen tekeminen ja opettaminen osoittautui helpoksi ja nopeaksi. Suurin työ on hyvän ja riittävän laajan opetusjoukon kerääminen. Kun sopiva aineisto on kerättynä, ennustemallin tekeminen ei ole ongelma. Ainoa vaatimus on riittävän tehokkaan tietokoneen löytäminen opetusta varten (nyt käytössä olivat CSC - Tieteellinen Laskenta Oy:n koneet). Opetetun neuroverkon käyttäminen ei enää vaadi suuria tehoja.

Liikennemääräennusteet onnistuivat tasaisemmin kuin nopeusennusteet. Vaikka kumulatiivisia summakäyriä tarkastellen nopeusennusteet olivatkin tarkempia, niissä oli systemaattisia virheitä. Nopeusennusteissa tulokset olivat epäluotettavia alimpien nopeuksien osalta ja malleilla oli taipumus yliarvioida näitä nopeuksia. Liikennemääräennusteissa esim. virhemarginaali, jonka sisään 90 prosenttia ennusteista mahtuu, oli laajempi kuin nopeusennusteissa, mutta mallit eivät tehneet systemaattisia virheitä.

Tutkimuksessa käytetyn aineiston (huonolla) laadulla ja määrän rajallisuudella on varmasti ollut oma vaikutuksensa tuloksiin. Karkeasti virheellisillä havainnoilla voi olla merkittävä vaikutus algoritmin toimintaan, sillä näiden havaintojen oppiminen huonontaa mallin yleistämiskykyä todennäköisempien syöteavaruuden alueiden kustannuksella (Schalkoff 1992). Käytetty datajoukko oli pieni, jolloin ruuhka- ja häiriötilanteiden aineiston määrä jäi vähäiseksi. Tämä johti mm. siihen, että mallin oli vaikea onnistua ruuhkanopeuksien ennustamisessa ja ettei ajankohtatieto parantanut mallia. Pieni aineisto johti myös pieneen piiloneuronien maksimimäärään eikä tällöin voitu kokeilla, kuinka paljon tulokset olisivat parantuneet yksinkertaisesti piiloneuronien määrää kasvattamalla. Myöskin kaistakohtaisten tietojen hyödyntäminen jäi selvittämättä aineiston pienen koon takia (kaistakohtaiset tiedot lisäävät syötteiden määrää). Tulokset olivat kuitenkin jo nyt varsin hyviä, mikä osoittaa menetelmän lupaavuuden.

12 Yhteenveto

Dynaamiset liikenneinformaatio- ja liikenteenohjausjärjestelmät eivät voi toimia reaktiivisesti, vaan niiden täytyy osata ennakoida liikennetilanteet ja sopeuttaa strategiansa ajoissa muuttuviin liikennetilanteisiin. Tämän tutkimuksen tarkoituksena oli selvittää eri tekijöiden vaikutuksia lyhyen aikavälin liikennetilanneennusteisiin.

Ennustemalleja tehtiin ensisijaisesti Länsiväylälle. Kehitettyjä menetelmiä testattiin kuitenkin myös kaksikaistaisessa kohteessa, valtatiellä 3 Valkeakoskella.

Ennusteet tehtiin MLP-neuroverkkojen avulla. Tutkimuksessa haluttiin ensin selvittää, minkälainen on lyhyen aikavälin liikennetilanne-ennusteen kannalta optimaalinen neuroverkko ja miten aineistoa kannattaa kerätä. Tämän jälkeen optimoitiin muita ennustemalliin liittyviä tekijöitä, kuten syötteitä, ennustejakson pituutta ja poikkileikkausten sijaintia. Lopuksi neuroverkkoon perustuvia malleja verrattiin naiiviin malliin ja Länsiväylällä kerättyjä kokemuksia sovellettiin kaksikaistaisessa kohteessa.

Parhaaksi aktivaatiofunktioyhdistelmäksi osoittautui piilokerroksen hyperbolinen tangentti yhdistettynä vastekerroksen lineaariseen funktioon. Aineiston esikäsitelymenetelmien välillä ei ollut suuria eroja. Skaalaamisen havaittiin useimmiten sopivan liikennemäärien ennustamiseen paremmin kuin normeerauksen ja pääkomponenttianalyysin, nopeusennusteessa tilanne oli päinvastainen.

Osamalleihin jaetun ennustemallin tulosten pitäisi olla vähintään yhtä hyviä kuin jakamattoman mallin. Tulos voitiin vahvistaa sellaisen mallin avulla, jossa liikennemäärä- ja nopeusennusteet tehtiin eri osamalleissa, mutta mallin, jossa jokaiselle ennustejaksolle oli oma osamalli, tulokset olivat sitä vastoin huonommat kuin jakamattoman mallin. Tähän oli kuitenkin todennäköisesti syynä sattuma. Neuroverkon opetukseen kannattaakin kiinnittää huomiota. Opetuksen onnistumista kannattaa myös tarkastella validointi- ja testidatan avulla, jotta voidaan varmistua, että neuroverkkoa voidaan soveltaa myös opetusdatan ulkopuolella ilman, että tulokset kärsivät.

Saadut tulokset osoittivat, etteivät kellonaika- ja päivätiedot parantaneet ennusteen laatua. Kellonaika- ja viikonpäivätiedon antotavalla ei näyttänyt olevan vaikutusta tuloksiin. Nopeuskeskihajontatiedot sitä vastoin paransivat etenkin nopeusennusteita. Kun tarkasteltiin regressioanalyysin tuloksia, voitiin havaita, että alhaisimpien nopeuksien osalta mallin antamat ennusteet olivat parempia kuin vertailumallin. Keskimääräisissä virhetermeissä eroa ei kuitenkaan huomattu.

Hypoteesina oli, että ennusteen tarkkuus huononee ennustejakson pituuden kasvaessa. Keskimäärin tarkasteltuna 30 minuutin ennuste olikin hieman epätarkempi kuin 15 minuutin vertailuennuste, mutta esimerkiksi suhteellisten virheiden itseisarvojen summakäyriä tarkastellessa ero oli häviävän pieni.

Kun poikkileikkauksia oli kaksi ja lisätietopoikkileikkaus sijaitsi ennustepoikkileikkauksesta ylävirtaan, ennuste oli sitä parempi, mitä kauempana poikkileikkaukset sijaitsivat toisistaan (kun puhutaan suhteellisen lyhyistä etäisyyksistä). Tilanne oli kuitenkin päinvastainen, jos lisätietopoikkileikkaus sijaitsi ennustepoikkileikkauksen jälkeen. Lisäksi havaittiin, että on olemassa etäisyys (tässä tapauksessa 800 metriä), jota lähemmäksi poikkileikkauksia ei kannattanut sijoittaa silloin, kun niiden määrä on rajallinen. On kuitenkin hyvä pitää mielessä, että poikkileikkausten optimaalinen sijainti on voimakkaasti riippuvainen kulloisestakin kohteesta, eli ettei nyt saatuja tuloksia voi välttämättä soveltaa muualla.

Neuroverkko malleille tehtiin vertailumalliksi ns. naiivi malli. Malli perustui liikenteen säännönmukaisuuteen ja sen tekemä ennuste pohjautui kunakin päivätyyppinä kullakin ajanhetkellä keskimääräisiin liikennemääriin ja keskinopeuksiin. Neuroverkkomallin antamat ennusteet olivat tarkempia kuin naiivin mallin, koska naiivi malli ei pysty reagoimaan odottamattomiin liikenteenhäiriöihin, jolleivät ne ole toistuvia. Liikennemääräennusteissa naiivi malli onnistui paremmin kuin nopeusennusteissa. Olennainen ero mallien välillä on kuitenkin se, että riittävällä aineistolla ja oikeilla syötteillä neuroverkkomalli on mahdollista opettaa reagoimaan satunnaisiin häiriötilanteisiin, mutta naiivissa mallissa tämä ei ole edes teoriassa mahdollista.

Työssä testattiin, tekeekö ruuhka-ajan aineistolla opetettu malli parempia ruuhka-ajan ennusteita kuin koko päivän aineistolla opetettu malli. Saadut tulokset olivat kuitenkin suunnilleen yhtä hyviä. Tilanne olisi saattanut olla toinen, jos opetusjoukko olisi ollut suurempi.

Kaksikaistaiselle tielle tehdyn ennustemallin perusteella havaittiin, että on tärkeämpää saada yhden suunnan liikennetiedot kahdesta pisteestä kuin yhdestä pisteestä molemmilta suunnilta. Vastaantulevan liikenteen tietojen vaikutus mallin suoritukseen oli suhteellisen pieni. Vaikuttaa siltä, että Länsiväylältä saatuja kokemuksia voidaan soveltaa suurelta osin suoraan myös kaksikaistaisilla teillä.

Ennustemallia suunniteltaessa mallille ei kannata antaa syöteparametreinä kaikkea mahdollista tietoa. Etenkin silloin, jos opetusdataa kerätään lyhyen aikaa, kannattaa poikkileikkausten määrä ja mukaan otettavat parametrit miettiä tarkkaan. Jos ennustemallia halutaan käyttää ympäri vuoden, sen olisi toimittava kaikenlaisissa olosuhteissa. Tällöin myös opetusdata olisi kerättävä ajanjaksolta, johon mahtuu kaikenlaisia sää- ja keliolosuhteita. Tällaisella aineistolla opetettaessa sää- ja kelityypistä kannattaisi tehdä syöteparametri.

Neuroverkkojen tekeminen ja opettaminen osoittautui helpoksi ja nopeaksi. Suurin työ on hyvän ja riittävän laajan opetusjoukon kerääminen. Kun sopiva aineisto on kerätty, ennustemallin tekeminen ei ole ongelma. Ainoa vaatimus on riittävän tehokkaan tietokoneen löytäminen opetusta varten. Opetetun neuroverkon käyttäminen ei enää vaadi suurta laskentatehoa.

Liikennemääräennusteet onnistuivat tasaisemmin kuin nopeusennusteet. Vaikka kumulatiivisia summakäyriä tarkastellen nopeusennusteet olivatkin tarkempia, niissä oli systemaattisia virheitä. Liikennemääräennusteissa virhemarginaali oli laajempi kuin nopeusennusteissa, mutta mallit eivät tehneet systemaattisia virheitä. Saatuja tuloksia voidaan pitää lupaavina.

Tutkimuksessa käytetyn aineiston (huonolla) laadulla ja määrän rajallisuudella on varmasti ollut oma vaikutuksensa tuloksiin. Karkeasti virheellisillä havainnoilla voi olla merkittävä vaikutus algoritmin toimintaan, sillä näiden havaintojen oppiminen huonontaa mallin yleistämiskykyä todennäköisempien syöteavaruuden alueiden kustannuksella (Schalkoff 1992). Käytetty datajoukko oli pieni, jolloin ruuhka- ja häiriötilanteiden aineiston määrä jäi vähäiseksi. Tämä johti mm. siihen, että mallin oli vaikea onnistua ruuhkanopeuksien ennustamisessa ja että piiloneuronien maksimimäärä jäi pieneksi. Myös kaistakohtaisten tietojen hyödyntäminen jäi tästä syystä selvittämättä. Tulokset olivat kuitenkin jo nyt varsin hyviä, mikä osoittaa menetelmän lupaavuuden.

Lähdeluettelo

ASHOK K, BEN-AKIVA M (1993). *Dynamic Origin-Destination Matrix Estimation and Prediction for Real-Time Traffic Management Systems*. 12th International Symposium on the Theory of Traffic Flow and Transportation, Berkeley, California. Transportation and Traffic Theory, Elsevier, Amsterdam, New York. s. 465 - 484.

BEN-AKIVA M, CANTARELLA G, CASCETTA E, RUITER J DE, WHITTAKER J, KROES E (1992). *Real-Time Prediction of Traffic Congestion*. The 3rd International Conference on Vehicle Navigation and Information Systems, Oslo, Norway. Conference Record of Papers, Institute of Electrical and Electronics Engineers, New York, N. Y. s. 557 - 562.

DEMUTH H, BEALE M (1998). *Neural Networks Toolbox for Use with Matlab*. User's Guide, Version 3. The Math Works Inc. s. 5-1 - 5-58.

HAYKIN S (1999). *Neural Networks*. A Comprehensive Foundation. International Edition, Second Edition. Prentice Hall International, Inc, USA. 842 s.

HOBEIKA A, OZBAY K (1991). *A Dynamic Traffic Assignment Model in the Context of Real Time Driver Information System*. 24th ISATA International Symposium on Automotive Technology and Automation, Florence, Italy, 20 - 24th May 1991. Automotive Automation Limited, Croyden, England. s. 473 - 480.

INNAMAA S (1999). *Automaattiset liikenteenohjaus- ja liikenneinformaatiojärjestelmät*. Tielaitoksen selvityksiä 28/1999, Tielaitos, Helsinki. 136 s.

INNAMAA S, VANHANEN K, PURSULA M (2000). *Länsiväylän automaattisen liikenteenohjausjärjestelmän vaikutukset liikennevirtaan*. Tielaitoksen selvityksiä 53/2000, Tielaitos, Helsinki. 82 + 8 s.

KARVONEN T, KETTUNEN J, VAKKILAINEN P (1982). *Kalman-suodattimen käyttö hydrologisten ja ekologisten mallien tilan identifioinnissa ja parametriestimoinnissa*. Monistesarja 1982:1, Teknillinen korkeakoulu, Vesitalouden laboratorio, Espoo. 45 s.

KWON E (1991). *A New Approach for Real-Time Prediction of Traffic Demand-Diversion in Freeway Corridors*. Applications of Advanced Technologies in Transportation Engineering: Proceedings of the Second International Conference, New York, American Society of Civil Engineers. s. 398 - 402.

KWON E, STEPHANEDES Y (1994). *Comparative Evaluation of Adaptive and Neural-Network Exit Demand Prediction for Freeway Control*. Transportation Research Record 1446. s. 66 - 76.

LEE S, KIM D, KIM J, CHO B (1998). *Comparison of Models for Predicting Short-Term Travel Speeds*. Konferenssi-cd, 5th World Congress on Intelligent Transport Systems, 12 – 16 October 1998, Seoul, Korea. 9 s.

MATSUI H, FUJITA M (1998). *Travel Time Prediction for Freeway Traffic Information by Neural Network Driven Fuzzy Reasoning*. Neural Networks in Transport Applications. (Toimittajat: Himanen V, Nijkamp P, Reggiani A.) s. 355 - 364.

OJA E (1998). *Neuraalilaskennan perusteet*. Luentomoniste, Teknillinen korkeakoulu, Informaatiotekniikan laboratorio. 199 s.

PAPAGEORGIOU M, BLOSSEVILLE J, HADJ-SALEM H (1990). *Modeling and Real-Time Control of Traffic Flow on the Southern Part of Boulevard Peripherique in Paris: Part I: Modeling*. Transportation Research A, Vol. 24A, No. 5, 1990. s. 345 - 359.

SCHALKOFF R (1992). *Pattern Recognition. Statistical, Structural and Neural Approaches*. John Wiley & Sons, Inc. 364 s.

SMITH B, DEMETSKY M (1994). *Short-Term Traffic Flow Prediction: Neural Network Approach*. Transportation Research Record 1453. s. 98 - 104.

SMITH B, DEMETSKY M (1997). *Traffic Flow Forecasting: Comparison of Modeling Approaches*. Journal of Transportation Engineering, Vol. 123, No. 4, July / August 1997. s. 261 - 266.

STEPHANEDES Y, ARGIROPOULOS I, MICHALOPOULOS P (1990a). *On-Line Traffic Assignment for Optimal Freeway Corridor Control*. Journal of Transportation Engineering, Volume 116, No. 6. s. 744 - 755.

STEPHANEDES Y, KWON E, MICHAPOULOS P (1990b). *On-Line Diversion Prediction for Dynamic Control and Vehicle Guidance in Freeway Corridors*. Transportation Research Record, No. 1287. s. 11 - 19.

STEPHANEDES Y (1991). *Dynamic Diversion Prediction for Real-Time Control in IVHS Networks*. 24th ISATA International Symposium on Automotive Technology and Automation, Florence, Italy, 20 - 24th May 1991. Automotive Automation Limited, Croyden, England. s. 459 - 463.

TIELAITOS (1999). *Liikennemääräkartta 1999*. Hämeen tiepiiri. Tielaitos, Tiehallinto.

ZHANG H, RITCHIE S, LO Z-P (1997). *Macroscopic Modeling of Freeway Traffic Using an Artificial Neural Network*. Transportation Research Record 1588. s. 110 - 119.

Käytetyt symbolit

b	Piste, jossa suora leikkaa y-akselin
β	Kerroin
C	Opetusjoukko
d	Toivottu vaste
e, \mathbf{e}	Virhe, virhematriisi
E	Hetkellinen kokonaisenergia
E_{av}	Keskimääräinen neliövirheen energia
φ	Aktivaatiefunktio
\mathbf{g}	Gradienttimatriisi
h	Piilokerroksen indeksi
\mathbf{H}	Hessen matriisi
i	Syötekerroksen indeksi
i, j	Yleinen neuroverkon kerroksen indeksi
\mathbf{J}	Jacobin matriisi
m	Syötteiden määrä, suoran kulmakerroin
μ	Skalaari
n	Opetuskierros
N	Neuronien lukumäärä kerroksessa
η	Oppimisaste
o	Vastekerroksen indeksi
q	Liikennemäärä
r	Korrelaatiokerroin
T	Opetusjoukon koko
u	Satunnaismuuttuja
v	Neuronista aktivaatiefunktioon menevä signaali, nopeus

w	Painokerroin
x, \mathbf{x}	Satunnaismuuttuja, satunnaismatriisi
y	Neuronin syöte/vaste

Pääkomponenttianalyysi

Merkitään \mathbf{X} :llä syöteavaruuden m -dimensioista satunnaisvektoria. Oletetaan, että \mathbf{X} :n oletusarvo on nolla ($E[\mathbf{X}] = \mathbf{0}$). (Jos näin ei ole, \mathbf{X} voidaan normeerata nollaodotusarvoiseksi). Merkitään \mathbf{q} :lla m -dimensioista ominaisvektoria, jolle vektori \mathbf{X} projisoidaan. Projektio A määritellään vektoreiden \mathbf{X} ja \mathbf{q} sisätulona.

$$A = \mathbf{X}^T \mathbf{q} = \mathbf{q}^T \mathbf{X} \quad (1)$$

Rajoitutaan tapauksiin, joissa \mathbf{q} :n normi on yksi ($\|\mathbf{q}\| = (\mathbf{q}^T \mathbf{q})^{1/2} = 1$). Projektio A on satunnaismuuttuja, jonka odotusarvo ja varianssi ovat riippuvaisia satunnaisvektorista \mathbf{X} . Jos satunnaisvektorin \mathbf{X} odotusarvo on nolla, voidaan osoittaa, että myös projektion A odotusarvo on nolla ja A :n varianssi on sama kuin sen keskimääräinen neliö.

$$E[A] = \mathbf{q}^T E[\mathbf{X}] = 0 \quad (2)$$

$$\sigma^2 = E[A^2] = E[(\mathbf{q}^T \mathbf{X})(\mathbf{X}^T \mathbf{q})] = \mathbf{q}^T E[\mathbf{X} \mathbf{X}^T] \mathbf{q} = \mathbf{q}^T \mathbf{R} \mathbf{q} \quad (3)$$

Yhtälössä $m \times m$ -matriisi \mathbf{R} on satunnaisvektorin \mathbf{X} korrelaatiomatriisi, joka on siis muodollisesti sama kuin odotusarvo \mathbf{X} :n ulkotulosta itsensä kanssa ($\mathbf{R} = E[\mathbf{X} \mathbf{X}^T]$). Havaitaan, että korrelaatiomatriisi \mathbf{R} on symmetrinen ($\mathbf{R}^T = \mathbf{R}$). Symmetrisyydestä seuraa, että jos \mathbf{a} ja \mathbf{b} ovat mitkä tahansa $m \times 1$ -vektorit, seuraava lauseke pätee.

$$\mathbf{a}^T \mathbf{R} \mathbf{b} = \mathbf{b}^T \mathbf{R} \mathbf{a} \quad (4)$$

Havaitaan, että edellä määritelty projektion A varianssi (3), on ominaisvektorin \mathbf{q} funktio. Näin ollen voidaan kirjoittaa

$$\sigma^2 = \mathbf{q}^T \mathbf{R} \mathbf{q} = \psi(\mathbf{q}) \quad (5)$$

Tämän yhtälön perusteella $\psi(\mathbf{q})$:ta voidaan pitää varianssin koettimena. (Haykin 1999.)

Seuraavana tehtävänä on löytää sellaiset ominaisvektorit \mathbf{q} , joilla $\psi(\mathbf{q})$ on paikallisessa minimissä tai maksimissa. Ratkaisu löytyy korrelaatiomatriisin \mathbf{R} ominaisrakenteesta. Jos \mathbf{q} on sellainen ominaisvektori, että $\psi(\mathbf{q})$ saa ääriarvon, tällöin mille tahansa ominaisvektorin \mathbf{q} pienelle häiriölle $\delta \mathbf{q}$ pätee

$$\psi(\mathbf{q} + \delta \mathbf{q}) = \psi(\mathbf{q}) \quad (6)$$

Nyt edellä esitetyn varianssin määritelmän (5) perusteella voidaan johtaa

$$\psi(\mathbf{q} + \delta \mathbf{q}) = (\mathbf{q} + \delta \mathbf{q})^T \mathbf{R} (\mathbf{q} + \delta \mathbf{q}) = \mathbf{q}^T \mathbf{R} \mathbf{q} + 2(\delta \mathbf{q})^T \mathbf{R} \mathbf{q} + (\delta \mathbf{q})^T \mathbf{R} \delta \mathbf{q} \quad (7)$$

Jättämällä viimeinen termi huomiotta, yhtälö (7) voidaan kirjoittaa muodossa

$$\psi(\mathbf{q} + \delta \mathbf{q}) = \mathbf{q}^T \mathbf{R} \mathbf{q} + 2(\delta \mathbf{q})^T \mathbf{R} \mathbf{q} = \psi(\mathbf{q}) + 2(\delta \mathbf{q})^T \mathbf{R} \mathbf{q} \quad (8)$$

Käyttämällä hyväksi yhtälöä (6) nähdään, että

$$(\delta \mathbf{q})^T \mathbf{R} \mathbf{q} = 0 \quad (9)$$

Tämä ei kuitenkaan päde \mathbf{q} :n kaikille häiriöille $\delta \mathbf{q}$, vaan rajoittuu ainoastaan sellaisiin häiriöihin, joille $(\mathbf{q} + \delta \mathbf{q})$:n normi on yksi ($\|\mathbf{q} + \delta \mathbf{q}\| = (\mathbf{q} + \delta \mathbf{q})^T (\mathbf{q} + \delta \mathbf{q}) = 1$). Nyt siis vaaditaan, että $(\delta \mathbf{q})^T \mathbf{q} = 0$. Tämä tarkoittaa sitä, että häiriön $\delta \mathbf{q}$ täytyy olla kohtisuora vektoriin \mathbf{q} nähden ja näin ollen ainoastaan muutokset \mathbf{q} :n suunnassa ovat sallittuja. (Haykin 1999.)

Ominaisvektorin \mathbf{q} elementit ovat dimensiottomia fyysisessä mielessä. Jotta seuraavan lausekkeen molemmat tekijät saadaan samadimensioisiksi kuin \mathbf{R} , tarvitaan skaalaustekijää λ .

$$\begin{aligned} (\delta \mathbf{q})^T \mathbf{R} \mathbf{q} - \lambda (\delta \mathbf{q})^T \mathbf{q} &= 0 \\ \Rightarrow (\delta \mathbf{q})^T (\mathbf{R} \mathbf{q} - \lambda \mathbf{q}) &= 0 \\ \Rightarrow \mathbf{R} \mathbf{q} &= \lambda \mathbf{q} \end{aligned} \quad (10)$$

Tämä yhtälö määrittelee ne ominaisvektorit \mathbf{q} , joissa varianssikoetin $\psi(\mathbf{q})$ saa ääriarvot. (Haykin 1999.)

Yhtälö (10) tunnetaan ominaisarvo-ongelmana. Ongelmalla on ei-triviaali ($\mathbf{q} \neq \mathbf{0}$) ratkaisu ainoastaan tietyillä λ :n arvoilla. Näitä kutsutaan korrelaatiomatriisin \mathbf{R} ominaisarvoiksi ja niihin liittyviä \mathbf{q} :n arvoja ominaisvektoreiksi. Oikeat, ei-negatiiviset ominaisarvot kuvaavat korrelaatiomatriisia. Kukin ominaisvektori on erilainen, jos ominaisarvot eroavat toisistaan. Merkitään $m \times m$ -matriisin \mathbf{R} ominaisarvoja seuraavalla tavalla: $\lambda_1, \lambda_2, \dots, \lambda_m$. Vastaavat ominaisvektorit ovat $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_m$. Nyt voidaan kirjoittaa

$$\mathbf{R} \mathbf{q}_j = \lambda_j \mathbf{q}_j, \quad j = 1, 2, \dots, m \quad (11)$$

Järjestetään ominaisarvot suuruusjärjestykseen siten, että on λ_1 suurin ja λ_m pienin. Rakennetaan ominaisvektoreista $m \times m$ -matriisi $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_j, \dots, \mathbf{q}_m]$. Nyt yhtälöjoukko (11) voidaan esittää yhtenä yhtälönä.

$$\mathbf{R} \mathbf{Q} = \mathbf{Q} \mathbf{\Lambda} \quad (13)$$

Yhtälössä $\mathbf{\Lambda}$ on diagonaalimatriisi, jonka matriisin \mathbf{R} ominaisarvot muodostavat: $\mathbf{\Lambda} = \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_j, \dots, \lambda_m]$. Matriisi \mathbf{Q} on ortogonaalinen matriisi siinä mielessä, että sen sarakkeiden vektorit täyttävät ortogonaalisuuden ehdon:

$$\mathbf{q}_i^T \mathbf{q}_j = \begin{cases} 1, & j = i \\ 0, & j \neq i \end{cases} \quad (14)$$

Yhtäläillä voidaan kirjoittaa

$$\mathbf{Q}^T \mathbf{Q} = \mathbf{I} \quad (15)$$

Tästä seuraa, että yhtälö (13) voidaan ratkaista muotoon

$$\begin{aligned} \mathbf{Q}^T \mathbf{R} \mathbf{Q} &= \mathbf{\Lambda} \\ \Rightarrow \mathbf{R} &= \sum_{i=1}^m \lambda_i \mathbf{q}_i \mathbf{q}_i^T \end{aligned} \quad (16)$$

Yhtälöitä (16) käytetään matriisin \mathbf{R} ominaishajotelman määrittelemiseksi. (Haykin 1999.)

Pääkomponenttianalyysi ja matriisin \mathbf{R} ominaishajotelma ovat yksi ja sama asia. Tämä yhtäsuuruus on seurausta yhtälöistä (5) ja (16), joista voidaan nähdä, että varianssikoettimet ja ominaisarvot ovat yhtä suuret ($\psi(\mathbf{q}_j) = \lambda_j$, $j = 1, 2, \dots, m$). Yhteenvetona voidaan mainita, että korrelaatiomatriisin \mathbf{R} ominaisvektorit, jotka kuuluvat nollaodotusarvoiselle satunnaisvektorille \mathbf{X} , määrittelevät ominaisvektorit \mathbf{q}_j , jotka esittävät pääsuuntia, joissa varianssikoettimilla $\psi(\mathbf{q}_j)$ on ääriarvot. Näihin ominaisvektoreihin liittyvät ominaisarvot määrittävät varianssikoettimien ääriarvot $\psi(\mathbf{u}_j)$. (Haykin 1999.)

Yhtälö (1) voidaan esittää muodossa

$$\mathbf{a}_j = \mathbf{q}_j^T \mathbf{x} = \mathbf{x}^T \mathbf{q}_j, \quad j = 1, 2, \dots, m \quad (17)$$

Yhtälössä \mathbf{a}_j on \mathbf{x} :n projektiio ominaisvektorin \mathbf{u}_j esittämään pääsuuntaan. Vektoreita \mathbf{a}_j kutsutaan pääkomponenteiksi ja niillä on samat fyysiset dimensiot kuin datavektorilla \mathbf{x} . Alkuperäinen datavektori \mathbf{x} voidaan esittää myös muodossa

$$\mathbf{x} = \mathbf{Q} \mathbf{a} = \sum_{j=1}^m \mathbf{a}_j \mathbf{q}_j \quad (18)$$

Pääkomponenttianalyysillä on käytännön merkitystä siinä, että se on tehokas menetelmä dimensioiden vähentämiseen. Data voidaan esittää tehokkaasti siten, että siitä karsitaan pois ne lineaarikombinaatiot, joilla on pienet varianssit, ja jätetään jäljelle ainoastaan ne, joiden varianssi on suuri. Merkitään korrelaatiomatriisin \mathbf{R} suurimpia ominaisarvoja $\lambda_1, \lambda_2, \dots, \lambda_l$ ($l \leq m$). Vektoria \mathbf{x} voidaan approksimoida katkaisemalla l termin jälkeen loppuosa yhtälöstä (18).

$$\hat{\mathbf{x}} = \sum_{j=1}^l \mathbf{a}_j \mathbf{q}_j = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_l] \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_l \end{bmatrix} \quad (19)$$

Kun tunnetaan alkuperäinen vektori \mathbf{x} , voidaan jäljelle jääneiden pääkomponenttien joukko laskea yhtälön (17) perusteella seuraavasti. (Haykin 1999.)

$$\begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_l \end{bmatrix} = \begin{bmatrix} \mathbf{q}_1^T \\ \mathbf{q}_2^T \\ \vdots \\ \mathbf{q}_l^T \end{bmatrix} \mathbf{x} \quad (20)$$

ISSN 0788-3722
ISBN 951-726-702-9
TIEL 3200640